Neural Comput & Applic (2019) 31:2523–2535
DOI 10.1007/s00521-017-3211-5

ORIGINAL ARTICLE

Simple speed estimators reproduce MT responses and identify strength of visual illusion

Daiki Nakamura[1] · Shunji Satoh[1]

Received: 3 October 2016 / Accepted: 4 October 2017 / Published online: 2 November 2017

Abstract Computational models of vision should not only be able to reproduce experimentally obtained results; such models should also be able to predict the input–output properties of vision. Conventional models of MT neurons are based on the concept of velocity filtering, as proposed by Simoncelli and Heeger (Vis Res 38(5):743–761, 1998). As this report describes, we provide another interpretation of the computational function of MT neurons. An MT neuron can be a simple speed estimator with an upper limitation for correct estimation. Subsequently, we assess whether the MT model can account for illusory perception of "rotating drift patterns," by which humans perceive illusory rotation (clockwise or counterclockwise rotation) depending on the background luminance. Moreover, to predict whether a pattern causes visual illusion, or not, we generate an enormous set of possible visual patterns as inputs to the MT model: $8^8 = 16{,}777{,}216$. Numerical quantities of model outputs obtained through a computer simulation for 8^8 inputs were used to estimate human illusory perception. Results of psychophysical experiments demonstrate that the model prediction is consistent with human perception.

Keywords MT · Visual illusion · Lucas–Kanade method · Computational model

1 Introduction

Selectivity has been a major topic of neuroscience and its related computational theory for many years. Many researchers have dedicated their efforts to discovering the X-selectivity in neurons of various visual areas by presenting visual inputs of various kinds. As an example of X, orientation selectivity was discovered in the primary visual cortex (V1) [1]. Most V1 neurons maximally respond to a particular orientation of lines or edges, but not to the orthogonal ones. Other examples are curvature selectivity of secondary visual cortex (V2) neurons [2], velocity selectivity of the middle temporal area (MT) neurons [3], and so on [4–6]. Evidence of discovering X relies on the unimodal response-curve function $f(X)$ of recorded neurons. If a response $f(X)$ is unimodal, taking its maximum value when $X = X_0$, then many researchers tend to infer that the recording neuron would prefer X_0, which is designated as *preferred* X. From the viewpoint of signal processing, we might conclude that such neurons would be band-pass filters against quantity X with its maximum gain at X_0. Many computational models are based on the computational interpretation of *preferred* X or X*-filtering* [7–9].

However, *preferred* X or X*-filtering* might not necessarily be the one and only interpretation for all cases. Given an opportunity for fresh interpretation, one might understand visual systems from a different aspect and might derive different models based on a new interpretation of neural properties.

The first objective of our research is to provide another interpretation of unimodal functions of MT neurons, which respond strongly to particular velocity, v_0, of moving visual stimuli. A simple speed estimator (a proposed MT neuron model in this article) also shows unimodal properties; the estimator based on the Lucas–Kanade method [10] is

✉ Daiki Nakamura
daiki@hi.is.uec.ac.jp

Shunji Satoh
shunji@uec.ac.jp

[1] Graduate School of Information Systems, The University of Electro-Communications, Tokyo 182-8585, Japan

designed so that its output $\hat{v} = f(v)$ is as equal to the actual velocity v as possible, performing similarly to a radar gun if $0 \le v < v_0$, where v_0 is not the preferred speed but the upper limitation for correct estimation. If the velocity of a moving stimulus exceeds the limitation, as $v_0 < v$, then the velocity estimator would fail to estimate the correct speed. Such velocity estimators will exhibit a unimodal property of $f(v)$ if output $\hat{v}$ converges to zero (no response) for overly fast v exceeding v_0.

The second objective of this article is to propose a new means of model evaluation. We will attempt to discover unknown illusory patterns through numerical simulation of the MT model. A computational model should not only (i) reproduce neural properties and (ii) provide computational meaning of the properties, but also (iii) contribute to the discovery of unknown matters including neural and perceptual properties of our visual system. The third requirement relates to evaluation of its generalization ability. For example, if we develop a visual model that sufficiently describes human perceptual properties, then we might distinguish between illusory patterns and non-illusory ones by observing outputs of the model using numerical simulations incorporating all possible input stimuli.

To evaluate model requirement (iii) described above, we particularly examine Fraser–Wilcox (FW)-type stimuli as depicted in Fig. 1 [11]. Humans perceive illusory rotation when the FW stimuli disappear [12]. For convenience, we designate the illusory rotation after the disappearance of FW stimuli as *drift illusion* hereinafter. The direction of illusory rotation depends on the background luminance of the afterimage [13]. Clockwise rotations are perceived when the background luminance is bright (white), but counterclockwise rotation is perceived with a dark (black) background. Assuming that the prior stimuli of Fig. 1 comprise eight kinds of luminance values in one period (a circular sector of 45°), and assuming the luminance value as represented by eight digits, the number of possible FW-type patterns is $8^8 = 16{,}777{,}216$. Psychological experiments using human subjects are unsuitable to classify the 16 million patterns into illusory ones or not. Almost 400 days would be necessary for one human subject to classify 16 million patterns if the subject were forced to judge within 2 s/pattern with no break. However, an accurate computational model can classify them in 4 days if the model takes only a 20 ms/pattern to calculate the output by computer simulation. The authors emphasize that we can use a computational model as an indefatigable virtual subject. Then, we should apply computational models to discover unknown matters.

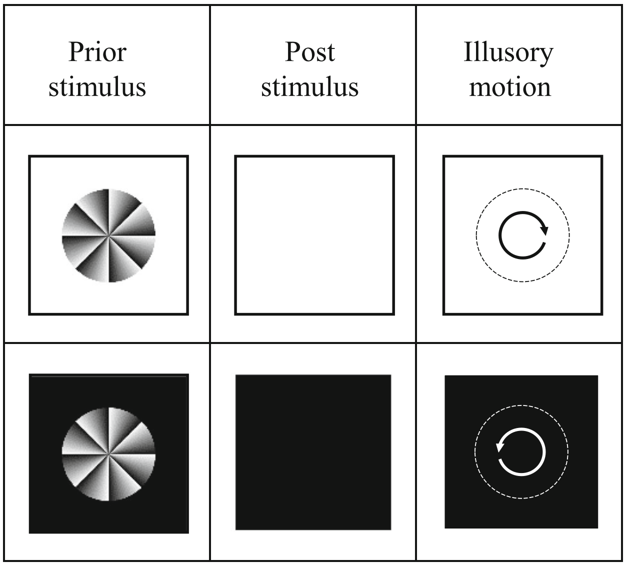

Fig. 1 Examples of drift illusion

As described in this paper, (1) we provide another interpretation of the computational function of MT neurons. An MT neuron can be a simple speed estimator with an upper limitation for correct estimation. Then, we develop an MT neuron model that is not based on brain science. Using the model, we examine whether the model reproduces MT responses of speed selectivity as presented in Fig. 2, or not, and (2) whether the model explains the luminance dependence of drift illusion, or not. Also, (3) we obtain model predictions for all possible patterns by numerical simulation. Additionally, we compare the model predictions to results obtained from psychological experimentation to evaluate the plausibility of our computational model.

2 Computational model of MT neurons

Simoncelli and Heeger [9] proposed an MT neuron model based on the concept of *velocity filtering*. An MT neuron of their model performs as a spatiotemporal filter defined in the frequency domain.

As another interpretation of the computational role of MT neurons, we propose the following idea: an MT neuron can be a simple velocity estimator for which the output $\hat{\boldsymbol{v}} = (\hat{v}_x \quad \hat{v}_y)^T$ (estimated velocity) is proportional to the true velocity $\boldsymbol{v}$ of visual inputs. We can derive such an estimator by minimizing an error function signified by $|\boldsymbol{v} - \hat{\boldsymbol{v}}|^2$, for which no concept of preferred velocity exists.

2.1 Basic computation and MT model

We propose that the Lucas–Kanade (LK) method [10], a computer vision algorithm for optical flow estimation minimizing a pre-defined error function, reveals the

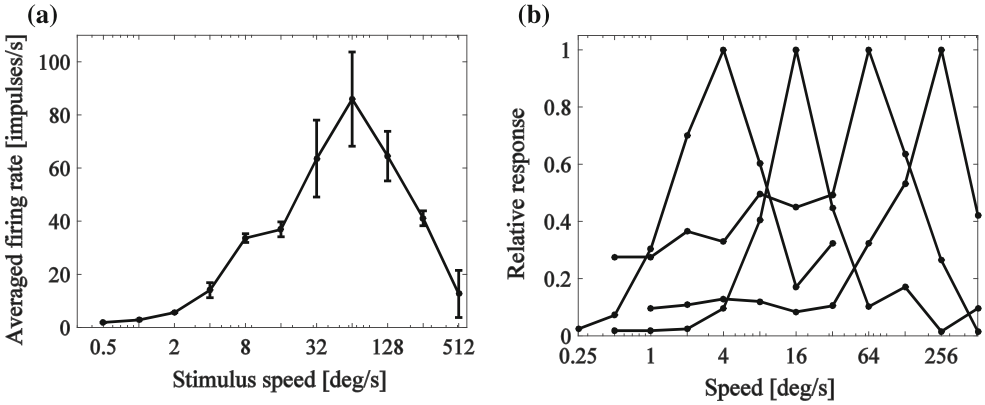

Fig. 2 Examples of MT neuron response [3]. **a** Solid lines show the firing rate of an MT neuron with respect to the speed of bar stimuli; bars represent the standard errors. **b** Relative responses of four MT neurons at different speeds

fundamental computation of MT cells. The LK method was derived under the following assumptions:

(a) Temporal changes of luminance are caused only by an objective motion.
(b) Spatial changes of luminance are approximated by the first-order Taylor expansion.
(c) Optical flows in a spatial window $w(x, y)$are constant.

The estimated velocity $\hat{\boldsymbol{v}}(x, y, t)$ calculated from the following equation minimizes the error between $\boldsymbol{v}$ and $\hat{\boldsymbol{v}}$ within a window $w(x, y)$.

$$\hat{\boldsymbol{v}}(x, y, t) = (-1)\left\{\begin{pmatrix} S_{xx}(x, y, t) & S_{xy}(x, y, t) \\ S_{xy}(x, y, t) & S_{yy}(x, y, t) \end{pmatrix} + \varepsilon^2 E\right\}^{-1} \times \begin{pmatrix} S_{xt}(x, y, t) \\ S_{yt}(x, y, t) \end{pmatrix} \quad (1)$$

$$S_{ij}(x, y, t) \overset{\text{def}}{=} w(x, y) * \left\{\frac{\partial I(x, y, t)}{\partial i}\frac{\partial I(x, y, t)}{\partial j}\right\} \quad (2)$$

$$(i, j = x,\ y,\ \text{or } t)$$

In that equation, $I(x, y, t)$ represents the relative luminance of the input image in the (x, y) spatial coordinate system at time t, where E represents an identity matrix, * denotes the convolution operator, and $\varepsilon^2 = 1.0 \times 10^{-4}$ is a scalar parameter that is applied to avoid division by zero. There is another implementation to avoid division by zero [14]. Also, $w(x, y)$ is a Gaussian window with standard deviation $\sigma_w = 11/6$ (window size is 11 × 11 pixels). The partial derivatives $\partial/\partial x$ and $\partial/\partial y$ for directional derivative are realized by numerical convolution between image I and the Gaussian derivative kernels of the spatial domain with size of $k \times k$ pixels [15–17]. The standard deviation of Gaussian derivative, σ_d, is proportional to kernel size k: $3\sigma_d = k/2$ pixel. The temporal derivatives $\partial/\partial t$ represent the difference of two adjacent frames: $I(t) - I(t-1)$. Speed estimation with various σ_d is equal to speed estimation with various spatial resolution. An estimator with a smaller σ_d (smaller k) provides a spatially higher-resolution map of optical flows. Although it is suitable for small objects, such an estimator cannot provide accurate estimation for rapid movements beyond its upper limitation. In contrast, a larger σ_d (larger k) is effective for rapid motion and for large objects at the sacrifice of spatial resolution. This tradeoff should be considered for accurate speed estimation.

Herein, we propose an alternative modeling concept of MT neurons: MT neurons are optical flow estimators; those neural outputs are proportional to the element (e.g., $\hat{v}_x$ or $\hat{v}_y$) of the estimated velocity. Apparently, the output of the LK method does not draw a unimodal profile, as presented in Fig. 2, because we do not base MT model on the concept of preferred speed. The output profile would be a monotonically increasing function with respect to input speeds. However, the LK method actually shows a unimodal profile.

Optical flows are estimated at all image positions of (x, y). We assume that an estimated optical flow parallel to the x-axis (zero degree, rightward motion), $\hat{v}_x$, is proportional to the neural activity (firing rate) of an MT cell selective to rightward (zero degree) motion of input. The spatial position (x, y) can be regarded as the receptive field center of the MT neuron, and σ_w, σ_d corresponds to the spatial region of receptive fields of the MT neuron. Although the details are presented in Sect. 2.3, changing the kernel size k, which is proportional to σ_d, can express a various peak speed of MT responses.

We generalize Eq. (1) for an arbitrary direction of vector components in addition to x (0°, horizontal) and y (90°, vertical) directions. Defining the local (ξ, η) coordinate system as the rotated (x, y)-system by degree ϕ, we obtain estimators for the ϕ and $\phi + 90°$ components of flows.

$$\begin{pmatrix}\hat{v}_\xi(\xi,\eta,t)\\ \hat{v}_\eta(\xi,\eta,t)\end{pmatrix} = (-1)\left\{\begin{pmatrix} S_{\xi\xi}(\xi,\eta,t) & S_{\xi\eta}(\xi,\eta,t)\\ S_{\xi\eta}(\xi,\eta,t) & S_{\eta\eta}(\xi,\eta,t)\end{pmatrix} + \varepsilon^2 E\right\}^{-1} \times \begin{pmatrix} S_{\xi t}(\xi,\eta,t)\\ S_{\eta t}(\xi,\eta,t)\end{pmatrix} \qquad (3)$$

The polar coordinate system is an example of (ξ,η)-system. $\hat{v}_\xi = \hat{v}_x$ and $\hat{v}_\eta = \hat{v}_y$ when $\phi = 0$. Partial derivatives with respect to i and j of Eq. (2), respectively, denote the directional derivatives along the ϕ and $\phi + 90°$ direction. Expanding Eq. (3), $\hat{v}_\xi$ is written as follows:

$$\hat{v}_\xi(\xi,\eta,t) = \frac{S_{\eta t}S_{\xi\eta} - S_{\xi t}\left(S_{\eta\eta} + \varepsilon^2\right)}{\left(S_{\xi\xi} + \varepsilon^2\right)\left(S_{\eta\eta} + \varepsilon^2\right) - S_{\xi\eta}^2} \qquad (4)$$

Assuming that $\hat{v}_\xi(\xi,\eta,t)$ is proportional to the neural activity of an MT cell that estimates the ϕ degree components of flows around (ξ,η), we formulate a new model of relative responses of MT neurons by normalizing Eq. (4). The following equation expresses the relative response of an MT model neuron estimating the ϕ degree component of flows.

$$\mathrm{MT}_\phi^{\mathrm{norm}}(\xi,\eta,t) = \frac{\hat{v}_\xi(\xi,\eta,t)}{\max_v \hat{v}_\xi(\xi,\eta,t)} = \frac{1}{L}\hat{v}_\xi(\xi,\eta,t) \qquad (5)$$

Therein, $L = \max_v \hat{v}_\xi(\xi,\eta,t)$ is introduced for normalization of neural activities [18]. We determine the constant value of L using moving random dots. Hereinafter, we designate the MT model based on the LK method (Eq. 4) as *the LK model*. Similarly, we designate the model of relative MT responses (Eq. 5) as *the normalized LK model*.

In the case of FW-type sequential inputs, the model estimation $\hat{v}$ is expected to be far from correct flows because the sudden disappearance of windmill object violates the assumption (a) of the LK method. In Sect. 3, we explore the consistency between the model outputs for FW-type stimuli and humans' illusory perception.

2.2 Numerical simulation

Using moving random dots, we simulated speed estimation by Eq. (4) with respect to the stimulus speed to ascertain whether the estimated speed presents a unimodal profile, or not. An input image composes Gaussian random dots of which each pixel value is drawn from the standard normal distribution. The image size was 150 × 150 pixels. Then, we prepared 20 sets of input images for each speed. The motion was limited in the x-axis direction (zero degree, rightward motion). Hereinafter, we show the temporal average of $\hat{v}_x(0,0,t)$, which is assumed to be proportional to the firing rate of an MT neuron with a receptive field on the center of images. The Gaussian derivative kernel size $k \times k$ was 5×5 pixels $(3\sigma_d = 5/2)$.

Figure 3 shows the averaged $\hat{v}_x$ for the rightward horizontal motion of inputs $v_x > 0$. Data on the left panel are identical to those of the right panel. The left panel is the linear plot for v_x, whereas the right panel is a semi-log plot. Dashed lines represent the standard errors. The ideal result of $\hat{v}_x$–v_x graph is $\hat{v}_x = v_x$ because Eq. (1) is designed just for correct estimation. When $v_x < 1$, we see $\hat{v}_x \simeq v_x$. However, estimated speeds $\hat{v}_x$ decrease gradually, when $v_x > 1$ pixel/frame, and eventually converge to zero.

Consequently, the speed or optical flow estimator based on the LK method shows a unimodal profile, but the algorithm is not based on the concept of preferred speed. Herein, we provide another interpretation of the speed taking the maximum response. It is not a preferred speed, but an upper limit for accurate estimation assuming that each MT neuron is a speed estimator.

2.3 Kernel size and MT profile

Figure 2b shows that different MT neurons possess different peak speeds. We show that such response curves emerge from setting a different kernel size k. Figure 4 portrays response curves for kernel sizes of four kinds based on the octave concept: $k_i = 2^{i+1} + 1,\ i = 1, 2, \ldots, N$. In this article, we set $N = 4$ and $k_1 = 5,\ k_2 = 9,\ k_3 = 17,\ k_4 = 33$ pixel. The left panel of Fig. 4 (linear plot) is averaged as $\hat{v}_x(k_i)$, whereas the right panel (semi-log plot) is averaged $\mathrm{MT}_{0°}^{\mathrm{norm}}(k_i)$. The normalizing factor L_i in $\mathrm{MT}_{0°}^{\mathrm{norm}}(k_i)$ is $L_i = \max_v \hat{v}_x(k_i)$, for example, $L_1 \simeq 1.1$, $L_2 \simeq 1.5$, as shown on the left panel of Fig. 4.

For simple notation, we hereinafter omit the coordinate variables and subscript of Eq. (4) or (5), e.g., $\hat{v}(k_i)$ and $\mathrm{MT}^{\mathrm{norm}}(k_i)$.

The results of Fig. 4 indicate that the larger kernel size pushes up the speed limitation. Observing the similarity between Figs. 2b and 4b, we can interpret the various profiles of MT neurons as speed estimators with different speed limitations.

In the case of $k_4 = 33$ (largest size of kernel), the speed was underestimated (Fig. 4a). The reason for this underestimation is a side effect of ε^2. The term of ε^2 becomes dominant for larger k because larger k causes smaller values of S_{xx}, S_{xy} and S_{yy} in Eq. (1). Future work shall include investigation of the kernel size dependence of ε^2 for correct estimation.

Selecting an appropriate kernel size in $k_i \in \{5, 9, 17, 33\}$ will be effective for correct speed estimation because of the tradeoff relation among different kernel sizes, written in Sect. 2.1. Kernel selection is discussed in Sect. 4.5.

Therefore, we propose the following computational interpretation of the MT population: the vision system

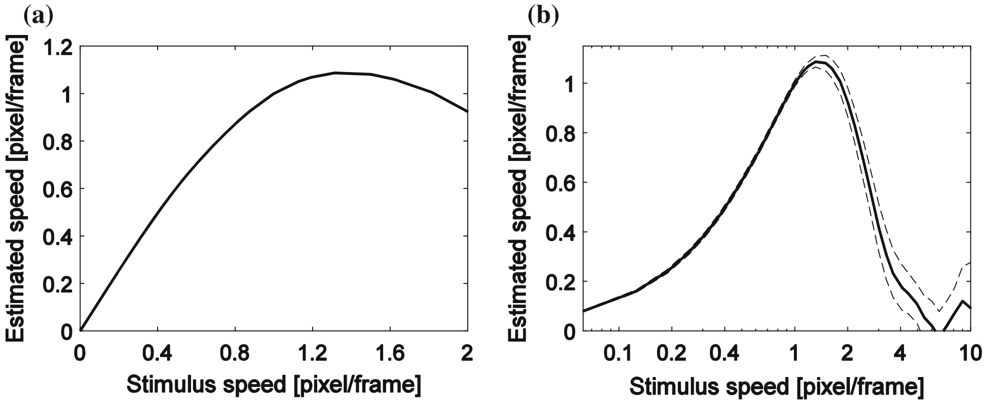

Fig. 3 Averaged speed of model-estimated speed $\hat{v}_x$ at different stimulus speeds: **a** the horizontal axis is linear; **b** the horizontal axis is logarithmic

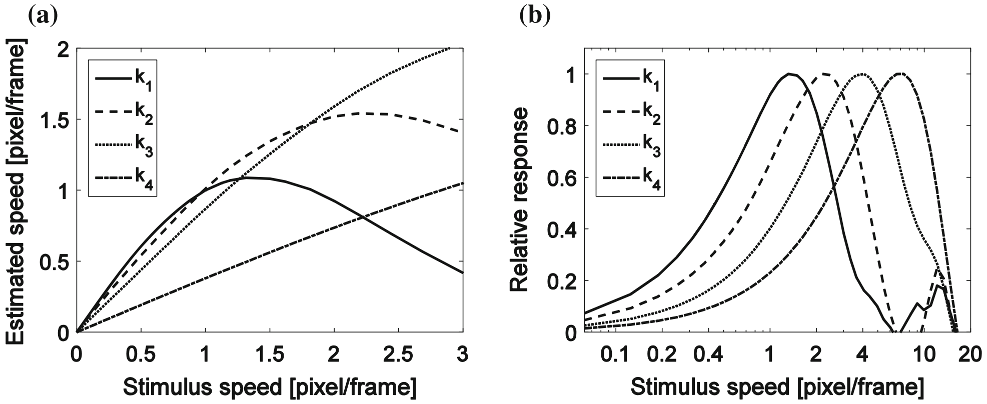

Fig. 4 Averaged estimated speeds obtained using various kernel sizes for calculating spatiotemporal derivative: **a** the horizontal axis is linear, and the **b** relative response normalized to its maximum value is shown on logarithmic scales for the horizontal axis

employs numerous MT neurons with different properties because of the tradeoff between spatial resolution and speed limitation.

2.4 Read-out from MT population

A read-out model from the outputs of MT population connects neural activity and motion perception. We derive a read-out model from our new interpretation of MT computation. The new concept is simple: every MT neuron tries to give its output proportional to the actual speed. Considering all speed estimators, $\hat{v}(k_i)$ are designed in accordance with the concept explained above. A simple method for speed estimation is averaging those outputs as follows:

$$\bar{v} = \frac{1}{N}\sum_{i=1}^{N} \hat{v}(k_i) \tag{6}$$

Rewriting Eq. (6) using Eq. (5), we deductively obtained the following read-out model for speed perception from MT populations.

$$\bar{v} = \frac{1}{N}\sum_{i=1}^{N} L_i \cdot \mathrm{MT}^{\mathrm{norm}}(k_i) \tag{7}$$

The model of Eq. (7) is coincidentally identical to that proposed by Boyraz and Treue [19], whose model accounts for the input-size effect on perceptual speeds. Section 5.1 presents discussion of the relation of Eq. (7), the Boyraz and Treue model, and other read-out models.

2.5 Discussion

Actually, MT neurons have been believed to tune for their preferred speed. However, response curves of the MT model based on the LK method, which is a simple speed estimator, also presents unimodal functions similar to MT response curves. The MT model was unable to estimate the stimulus speed correctly because exceeding a specific speed is a violation of assumption (b) of the LK method: “a change in luminance can be expressed by first-order approximation of the Taylor expansion.” Therefore, we can rephrase the statement of *preferred speed* by *upper limitation* of correct estimation.

Our examination revealed that the speed estimator based on the LK method also reproduced that MT neurons reached its maximum firing rate at various speeds, as portrayed in Fig. 4, using various kernel sizes for calculating the spatial derivative. This result demonstrates that each MT neuron estimates an optical flow with various kernel sizes. It is possible for the normalized LK model (Eq. 5) to be constructed as a neural network using V1 neuron models that calculate spatiotemporal derivative [15–17].

We reproduced the unimodal profile of MT outputs with respect to input stimulus. However, we recognize that the current model is insufficient to explain complex properties of MT neurons, e.g., contrast dependency [20], spatial frequency dependency [21], and texture dependency [22]. Those topics are left as subjects for future work.

3 Reproducing rotational illusion dependent on background luminance

We next assess the plausibility of our read-out model (Eq. 6 or 7) as a model of motion perception by comparing humans' response and model outputs using Fraser–Wilcox type stimuli, as depicted in Fig. 1. We expect that the estimated motion vectors (optical flows) are spatially rotating and that the direction of rotation depends on the background luminance.

3.1 Numerical simulation: rotational directions and the rotational strength

The input image size was 500 × 500 pixels. The circular pattern diameter was 300 pixels. Inputs are grayscale images of which luminance values were 0.0 (darkest, black) to 1.0 (brightest, white). Figure 5 presents the estimated optical flow vectors obtained using Eq. (6) ($k_i \in \{5, 9, 17, 33\}$). In Fig. 5, clockwise rotation vectors appeared when the relative background luminance was 1.0 (Fig. 5, top). In contrast, counterclockwise rotation vectors appeared when the relative background luminance of 0.0 (Fig. 5, bottom). Those results are qualitatively consistent with illusory perception by human subjects [13].

To evaluate our read-out model quantitatively, we define spatially averaged rotation $\bar{R}$ by the following formula, known as the rot operator introduced in vector analysis.

$$\begin{aligned}\bar{R} &= \frac{1}{|S|}\iint_S \mathrm{rot}_{2\mathrm{D}}\bar{\boldsymbol{v}}(x,y,t)\mathrm{d}S \\ &= \frac{1}{|S|}\iint_S \frac{\partial \bar{v}_y(x,y,t)}{\partial x} - \frac{\partial \bar{v}_x(x,y,t)}{\partial y}\mathrm{d}S \end{aligned} \quad (8)$$

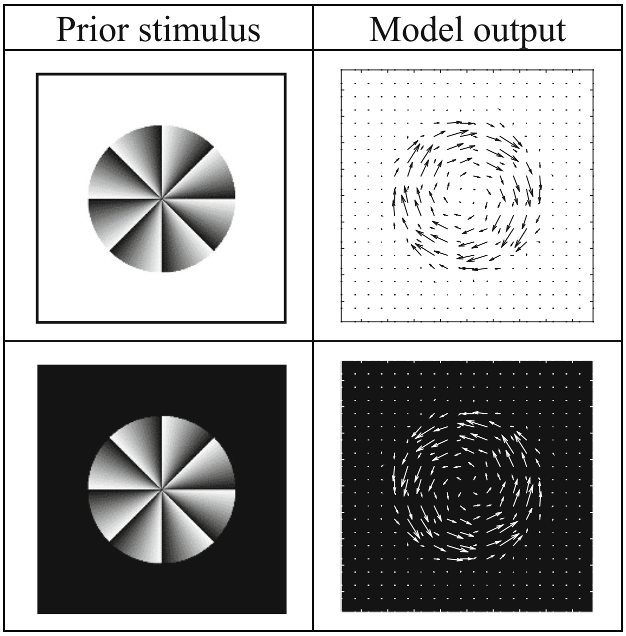

Fig. 5 Output vectors (optical flow, estimated perception of motion) obtained from our read-out model (Eq. 6)

Therein, S denotes the area of circular patterns. $\bar{R} > 0$ is associated with counterclockwise rotation, whereas $\bar{R} < 0$ coincides is associated with clockwise rotation. Figure 6 portrays the rotation $\bar{R}$ obtained from our read-out model with respect to background luminance. The smallest negative value of $\bar{R}$, clockwise rotation, was obtained at maximum background luminance ($I = 1.0$). In contrast, the largest positive value for counterclockwise rotation was obtained at minimum relative luminance ($I = 0.0$). The magnitude of rotation was zero at background luminance $I = 0.5$.

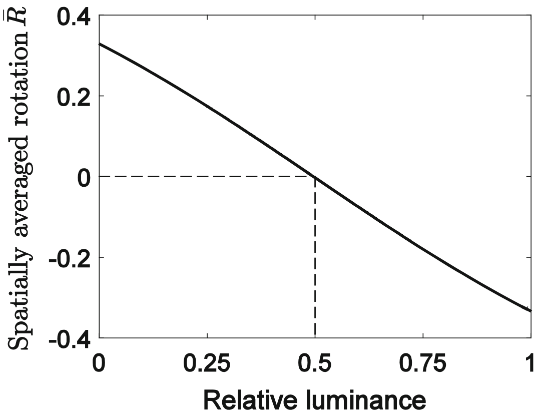

Fig. 6 Rotations of model outputs with respect to the relative luminance

3.2 Discussion

Results presented in the previous section indicate that the model accounts for human illusory perception for the drift illusion depending on the background luminance. The model includes the assumption that "(a) temporal changes of a texture are caused only by an objective motion." In other words, it does not presume suddenly disappearing objects such as in the case of the drift illusion. Although our read-out model's outputs for drift illusion are meaningless from an engineering perspective, it is interesting that these rotating vectors representing optical flows are consistent with human perception.

Theoretical reasons for the luminance dependence of illusory rotation can be considered. From Eq. (1), we ascertained that the temporal derivative term $\partial I/\partial t$ affects the rotational direction and rotational strength. For simplicity, we analyzed the illusion on the polar coordinate system (r, θ) using the center of FW stimuli as the origin (Fig. 7a). The right panel of Fig. 7 presents the relative luminance $I(r, \theta)$ of FW stimuli with respect to angle θ. The direction of optical flows is an almost angular direction. We restrict ourselves to considering the case of $\theta = 0 \sim 45^\circ$ because FW stimuli comprise periodic circular sectors of 45°. In the case of left panel of Fig. 7 ($\phi = \theta$), Eq. (3) is rewritten as

$$\begin{pmatrix} \hat{v}_\xi(r,\theta,t) \\ \hat{v}_\eta(r,\theta,t) \end{pmatrix} = (-1)\left(\frac{\partial I(r,\theta,t)}{\partial t} \Big/ \left\{ \begin{matrix} 0 \\ \frac{1}{r}\frac{\partial I(r,\theta,t)}{\partial \theta} \end{matrix} \right\} \right). \tag{9}$$

Herein, the luminance change of radial direction is zero ($\partial I/\partial \xi = 0$) and the window is the Dirac delta function ($w(r,\theta) = \delta(r,\theta)$). The parameter of avoiding zero division is zero ($\varepsilon^2 = 0$). From Eq. (9), the estimated angular velocity $\omega(r,\theta,t)$ is calculated as

$$\omega(r,\theta,t) = \frac{1}{r}\hat{v}_\eta(r,\theta,t) = -\frac{\partial I(r,\theta,t)}{\partial t} \Big/ \frac{\partial I(r,\theta,t)}{\partial \theta}. \tag{10}$$

Equation (10) shows that the sign of temporal change of luminance (numerator) and the sign of spatial change (denominator) determine the direction of rotation. Figure 8 portrays $\omega(\theta,t)$, $\partial I(\theta,t)/\partial t$, and $\partial I(\theta,t)/\partial \theta$ under background luminance of 0.0 (black) and 1.0 (white). Comparing the two columns in Fig. 8, the temporal derivative term causes rotational direction and rotational strength of drift illusion's dependency on background luminance. Only the temporal derivative term is dependent on background luminance.

The success of those analyses is attributable to application of a simple formula composed of spatiotemporal derivatives to the fundamental computation of MT cells. From results of psychophysical experiments using visual inputs similar to FW stimuli, Hsieh et al. [12] concluded that illusory motion might be related to the afterimage. As an alternative explanation of visual illusion for the FW stimuli, we demonstrated that the illusion might result from incorrect estimation of optical flows by MT neurons (Fig. 5). In this simulation, illusory optical flows related to illusory rotation appeared on just one frame because the temporal derivatives $\partial/\partial t$ were implemented by the difference of two adjacent frames: $I(t) - I(t-1)$: illusory motion appears immediately after disappearing visual stimuli. Moreover, the duration of illusory motion by the model is the frame interval. Actually, Hsieh et al. inferred that motion illusion lasts for a shorter time than the decay of the afterimage, and that the illusion observed only at the beginning phase of disappearing the FW-type stimuli. When $\partial/\partial t$ of the LK model is implemented by the temporal convolution kernel formulated by the Gaussian derivative with standard deviation σ_t [15–17], the duration of illusory optical flows is proportional to σ_t. Therefore, the

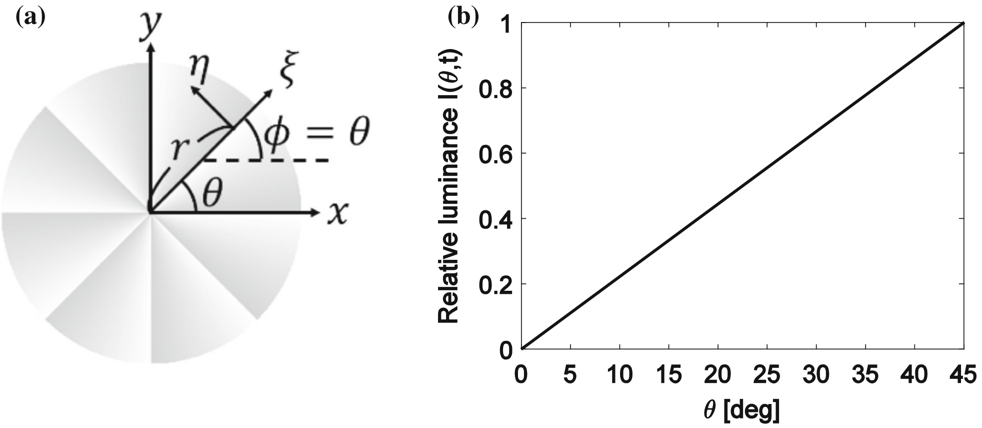

Fig. 7 Relative luminance of FW stimuli on the polar coordinate system $I(\theta)$: **a** FW stimulus and ξ, η axis and **b** relative luminance of FW stimulus with respect to the polar angle θ

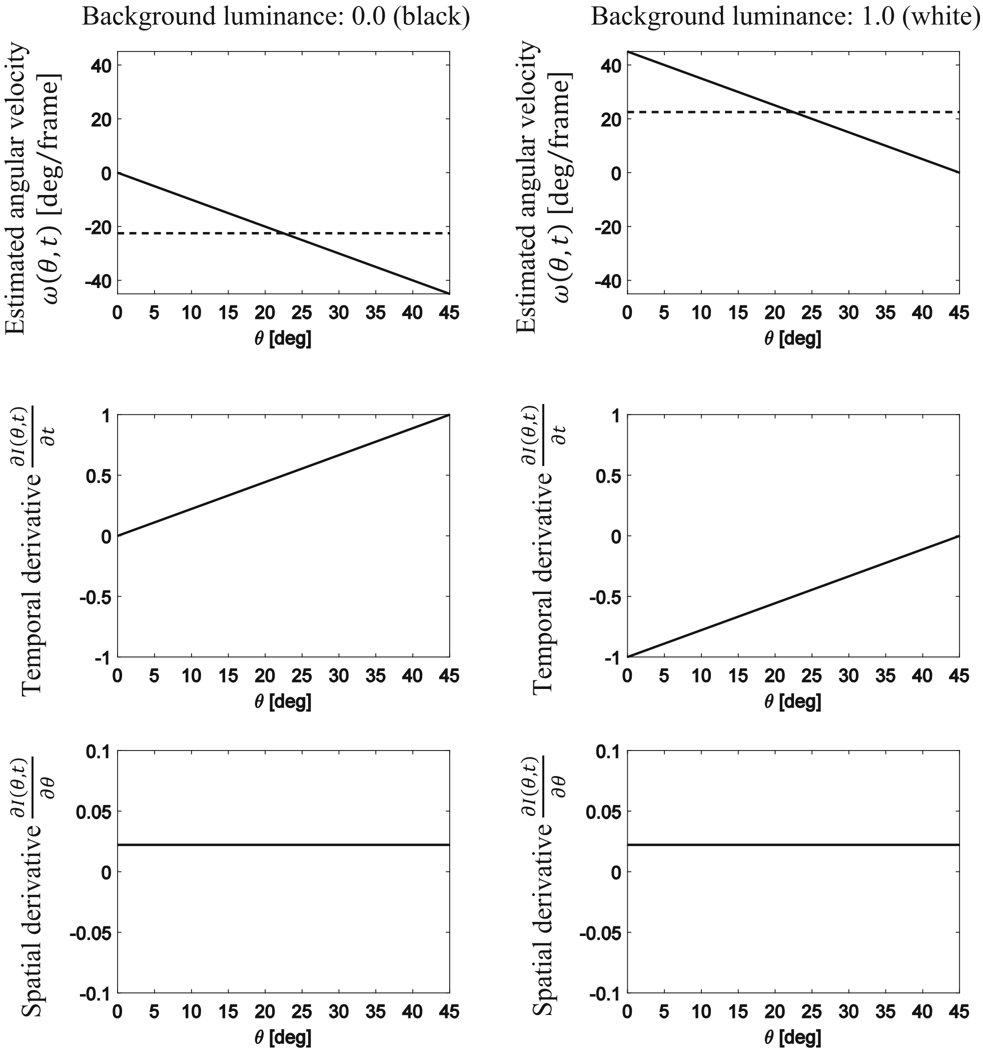

Fig. 8 Cause of drift illusion dependence on background luminance

model properties are consistent with the conjectures presented by Hsieh et al.

4 Model predictions and psychological experiments

Correlation between human perception and model prediction can be investigated using prior/post-images with white background to assess the MT model generalizability.

4.1 Circular stimulus

As portrayed in Fig. 9, a prior stimulus comprises circular sectors of 45°. A luminance pattern in a circular sector comprises eight gray levels: a combination of $I \in \{0/7, 1/7, 2/7, \ldots, 7/7\}$. The number of possible patterns is $8^8 = 16{,}777{,}216$.

4.2 Selection of stimuli for psychological experiment

We obtain rotation $\bar{R}$ for stimuli of over 16 million kinds on a white background. To reduce the simulation time, we set $N = 1$ and $k = 5$ pixel in a read-out model. We discuss model predictions using other kernels in Sect. 4.5. Figure 10 presents a histogram of rotation $\bar{R}$, indicating that almost all stimuli have small rotation $|\bar{R}| \simeq 0$, although some stimuli cause a clockwise or counterclockwise rotation vector. This result implies that almost no stimuli would be illusory patterns, but some patterns with large $|\bar{R}|$ might cause illusions for humans. The simulation time was less than 94 h (dual processor Xeon E5-2630 v2 2.6 GHz; Intel Corp.).

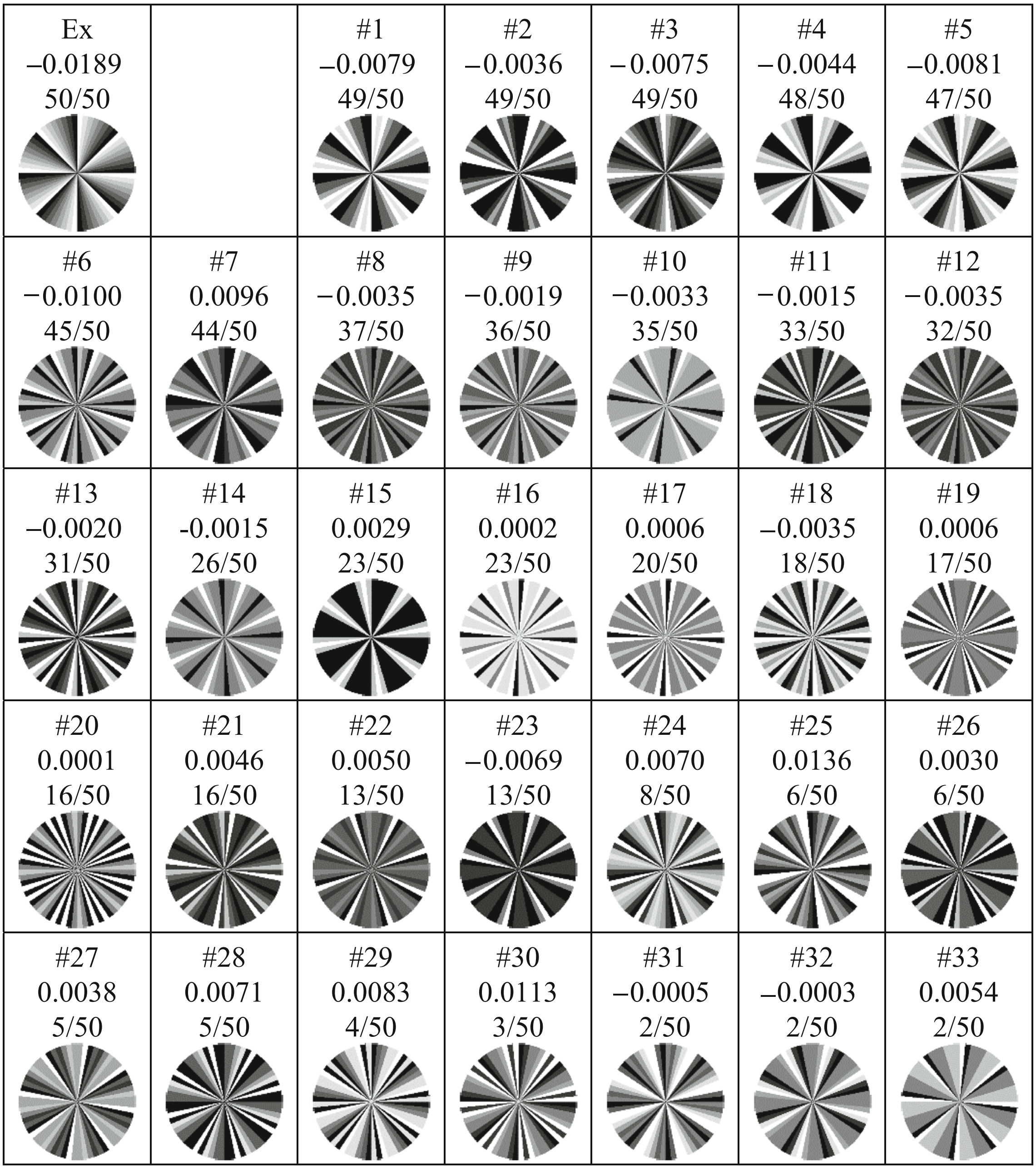

Fig. 9 Stimuli used in psychological experiments sorted by the probability of human judgment. #1–#33 are the indexes of stimuli; Ex. is FW stimulus drawn with eight grayscale level. Negative and positive real values are spatially averaged rotation $\bar{R}$. Fractional numbers represent the probability of human judgment to clockwise rotation of perception for 50 trials

For psychological experiments, 33 patterns were chosen randomly from 16 million patterns so that model predictions $\bar{R}$ were distributed uniformly, and so that a selected pattern contains both black and white ($I = 0.0$ and $I = 1.0$). Real values of Fig. 9 signify $\bar{R}$s from −0.0100 to 0.0136.

4.3 Methods

Each human subject was seated in a dark room with the head resting on a chin-rest fixed 1 m from the display. At the center of a gamma-corrected CRT monitor with a refresh rate of 85 Hz (GDM-F520; Sony Corp.), 33 selected stimuli were displayed. The display resolution was 1024 × 768 pixels. The screen visual angle was 22.0 × 16.6°. The circular stimulus diameter was 13.0° (300 pixels). The maximum luminance (white; $I = 1.0$) was 81.3 cd/m^2.

The 33 prior stimuli in Fig. 9 are displayed randomly. Each stimulus was displayed 10 times. Post-stimuli were uniformly white. Prior stimuli were presented for 1500 ms.

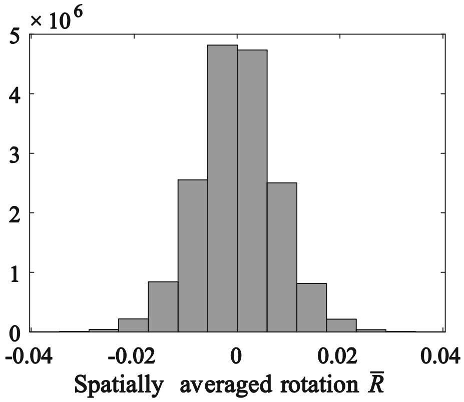

Fig. 10 Histogram of spatially averaged rotation $\bar{R}$ for 16,777,216 stimuli

Subsequently, prior stimuli disappeared; post-stimuli (uniform white) were displayed. Then, subjects were forced to report, as soon as possible, the direction of rotation after the disappearance prior stimuli (either clockwise or counterclockwise; 2AFC) displayed with a rotary device (PowerMate NA16029; Griffin Technology). The participants were five naïve subjects (23–24 years old). This study was approved by the ethics committee of the University of Electro-Communications.

4.4 Correlation between model output and human perception

The fractional number in Fig. 9 is the probability of human judgment for "clockwise" rotation for 50 trials. For example, 49/50 of #1 means that humans tend to perceive clockwise illusory rotation and 2/50 of #33 perceive counterclockwise rotating illusion.

Next, we compared model predictions with human responses. We adopt the following formula to transform rotation $\bar{R}$ into the stochastic judgment of clockwise motion $\Pr(CW)$.

$$\Pr(CW) = \frac{1}{2}\left(1 - \operatorname{erf}\left(\frac{\bar{R}}{s\sqrt{2}}\right)\right) \tag{11}$$

Therein, erf() is the error function; s is a positive parameter. We assumed that the chance level corresponds to circumstances in which $\bar{R} = 0$ and $\Pr(CW) = 0.5$. The free parameter s of Eq. (11) was determined by application of a nonlinear fitting of the model function $\Pr(CW)$ to 33 human judgment data. The best parameter was $s = 0.013$.

Figure 11 presents a scatter plot of model judgment and human judgment. The open circle corresponds to results for FW stimuli drawn with eight grayscale levels. Real values r and p in the upper left of Fig. 11 are Pearson's correlation coefficient and p value for testing the hypothesis of no correlation. If the model prediction were perfectly correct, then markers in Fig. 11 would be arranged on the diagonal line. The computational prediction of human perception was not perfect, but a positive correlation between them might be readily apparent (0.81 of correlation coefficient and $p < 10^{-8}$ for no-correlation testing). We obtained illusory patterns aside from the FW pattern, as shown in #1 and #33 of Fig. 9.

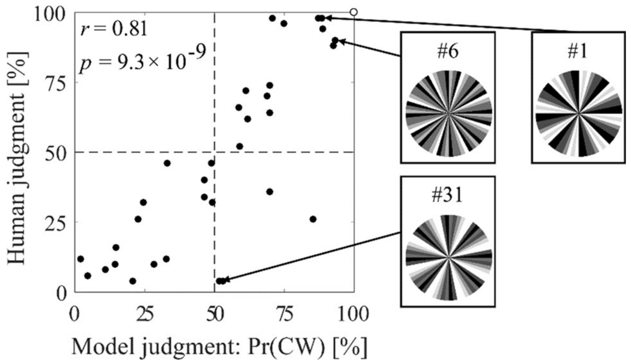

Fig. 11 Scatter plot of model judgment and human judgment with $N = 1$ and $k = 5$. An open circle at the upper right corner of the plot corresponds to the original FW stimulus drawn with eight grayscale levels, r is Pearson's correlation coefficient, and p is the p value for testing the no-correlation hypothesis

4.5 Effects of kernel size on model prediction

In the previous section, $N = 1$ (single kernel size) and $k = 5$ were set in the read-out model for simple simulation and discussion. In this section, we perform simulation with other parameter settings as follows: (i) $N = 1$ and $k \in \{5, 9, 17, 33\}$ and (ii) $N = 4$ (multiple kernel size) using all possible kernels of $\{5, 9, 17, 33\}$. We then evaluate correlation coefficients between model judgment and human judgment. Additionally, we investigate the effects of image size on correlation coefficients.

Solid lines of Fig. 12 show correlation coefficients r between the model judgment and human judgment with respect to kernel size k. Dashed lines are the correlation coefficient in the case of $N = 4$. Input images were scaled to obtain different image sizes using scale factors $f \in \{1/4, 1/2, 1/1\}$. From Fig. 12, it is apparent that a larger kernel size is better for larger input image. When $f = 1/2$, $N = 1$, and $k = 17$, the best correlation coefficient of 0.96 is obtained.

Figure 13 presents a scatter plot of model judgment and human judgment using the best parameters. Comparison of Figs. 11 and 13 shows improvement of the r value. These results indicate that a kernel size selection according to image size is an important computation accounting for visual perception.

Using multiple kernel size and the read-out by Eq. (6) scores a better r value, on average. Average r of multiple

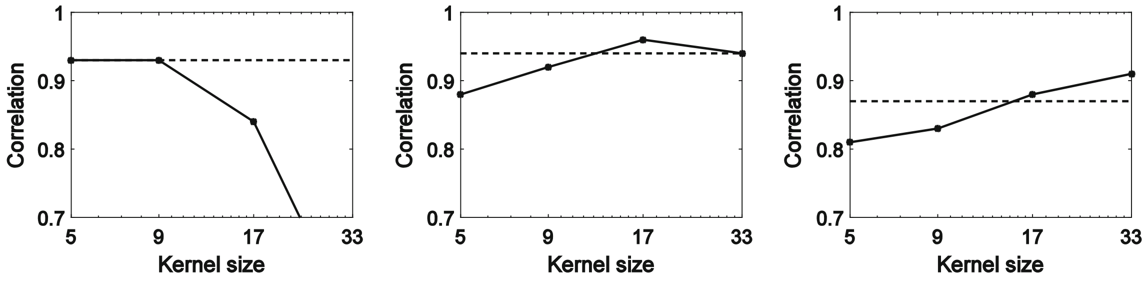

Fig. 12 Correlation coefficients between human judgment and model judgment using single (solid line) and multiple (dashed line) kernels with the resolution factor $f = 1/4$ (left), $f = 1/2$ (center), $f = 1/1$ (right, original scale)

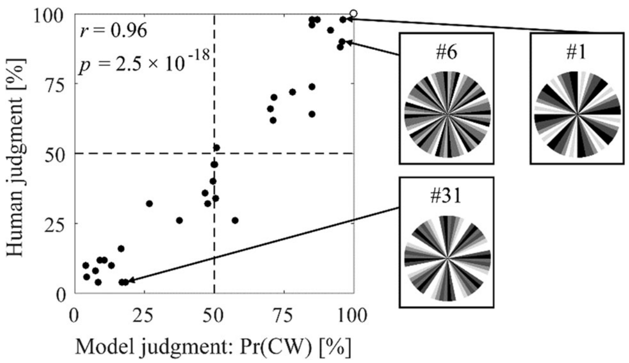

Fig. 13 Scatter plot of model judgment and human judgment of the best parameter ($N = 1$, $k = 17$, and $f = 1/2$)

kernels and single kernel were, respectively, 0.913 and 0.864. Object sizes and the best kernel sizes are factors that are unknown in advance. Therefore, a model using multiple kernels is expected to be useful in general cases to achieve rapid estimation.

5 General discussion and conclusion

We demonstrated that the response curves of the MT model based on the Lucas–Kanade method also show unimodal functions with respect to stimulus speed such as MT neuron response curves, although the model was not formulated to show unimodal responses. Our read-out model from MT population accounted for human illusory perception. First, in this section, we evaluate the model by comparison with the other characteristics of physiology and by comparison with other computational model of MT neurons. Second, we present clues to discover the novel illusory patterns.

5.1 MT model and read-out model

The tuning width, which is a full width at half maximum of tuning curve, is another aspect to evaluate the MT model plausibility. Maunsell and van Essen [3] reported that the average tuning width for speed of MT neurons is approximately a 7.7-fold change of speed (2.9 octaves). The tuning width of the normalized LK model $\mathrm{MT}_{0^\circ}^{\mathrm{norm}}(k = 5)$ (the smallest k; Fig. 4) is a 6.4-fold change of speed (2.7 octaves). The tuning width similarity is expected to support the plausibility of the LK model.

Assuming that MT neurons are velocity estimators, we obtained another interpretation of the peak speed of the MT tuning curve. It does not mean a *preferred* speed but an *upper limitation* for correct estimation of speed. Herein, we try to present a computational explanation of complex responses of MTs depending on the stimulus properties. Krekelberg et al. [20] reported that the peak speed of MT neurons decreased with lower contrast of displayed stimulus. This phenomenon is expected to be trivial because a lower-contrast input causes a lower signal-to-noise (SN) ratio. Consequently, the upper limitation for correct estimation also decreases for signals with a lower SN ratio. The side effect of parameter ε^2 of Eq. (1) is also related to the contrast dependence of peak speeds.

Boyraz and Treue [19] discovered that the peak speed of MT neurons becomes slower for smaller stimuli, which suggests that the smaller stimuli pushed down the upper limitation of collect speed estimation. Overly small stimuli violate assumption (c) of the LK method: optical flows in a spatial window $w(x, y)$ are constant. Future works must include an examination of whether the LK model reproduces the dependence on stimulus properties.

We compared our read-out model from the MT population (Eq. 7) with a modified labeled line model proposed by Boyraz and Treue [19] and vector averaging (center of mass). All of them share the same form.

$$v_p = \frac{\sum_{i=1}^{N}\left(\mathrm{MT}_i^{\mathrm{norm}} \times L_i\right)}{\alpha} \qquad (12)$$

Herein, v_p stands for the perceived speed (result of read-out from MT population), N signifies the number of MT neurons, $\mathrm{MT}_i^{\mathrm{norm}}$ denotes the relative response of an MT neuron, L_i represents a specific value (usually designated as "label") with a specific MT neuron, and α is a normalizing factor. Changing normalizing factor α in Eq. (12), we can express the three models: $\alpha = N$ corresponds to our read-out model, and $\alpha = \mathrm{const.}$ corresponds to Boryaz and Treue model. The original vector averaging model is given as $\alpha = \sum_{i=1}^{n} \mathrm{MT}_i^{\mathrm{norm}}$. Boyraz and Treue did not describe the computational meaning of introducing a constant α. Their model (constant α) reproduced misperceptions of speed perception dependent on the stimulus size. It is noteworthy that the model of constant α by Boyraz and Treue is computationally equivalent to our simple read-out model of Eq. (7), averaging estimated speeds, in which $\alpha = N$ is also a constant value. Therefore, the size dependence of motion perception can also be interpreted as a side effect of our read-out model.

5.2 Illusory motion perception

We obtained model predictions for all possible patterns by numerical simulation using our read-out model, which demonstrated strong positive correlation between human perceptions and model predictions. Unfortunately, we did not discover truly novel illusory patterns that are not FW-type stimuli. To reduce the simulation time, we limited prior stimuli to circular patterns, which include luminance values of eight kinds in one period. Some room exists for discovering new illusory patterns, although a quite longer simulation time will be necessary because the number of all possible two-dimensional patterns is $(m \times n)^d$. Herein, the size of input images is $m \times n$ pixels, with discretization of luminance by d levels.

Drift illusion causes illusory rotation to violate assumption (a) of the LK method: temporal changes of luminance are caused only by an objective motion. Therefore, the other assumptions (b) and (c) can serve as clues to discover new illusory patterns. For example, the roof edge violates assumption (b): spatial changes of luminance are approximated by the first-order Taylor expansion. Overly small stimuli or chaotic local motion also violates assumption (c): optical flows in a spatial window $w(x, y)$ are constant. Discovering completely novel illusory patterns based on those clues is left as a subject for future work.

5.3 Conclusions

First, we demonstrated that response curves of MT model based on the Lucas–Kanade method, a computer vision algorithm for optical flow estimation, also represent unimodal functions such as response curves of MT neurons. The peak speed at which an MT neuron reaches its maximum firing rate, usually called the preferred speed, can be inferred as an upper limit of correct speed estimation. Second, we demonstrated that our read-out model from MT population reproduced rotational illusion dependent on background luminance. Then, we sought to discover illusion patterns aside from well-known patterns. Numerical simulations exhibited strong positive correlation between human perception and model prediction.

Results of this study can elucidate visual systems from various aspects, facilitate the evaluation of various vision models, and facilitate generation of new illusory patterns.

Several other variations of the LK method exist [23]. Actually, MT models based on other methods might also reproduce MT responses and human motion perceptions. A subject left for future work is to distinguish which algorithm is the most suitable for use with the MT model.

Acknowledgements This work was partially supported by JSPS KAKENHI (24500371 and 16K00204) and NIJC Riken, Japan.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Hubel D, Wiesel T (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160:106–154
2. Ito M, Komatsu H (2004) Representation of angles embedded within contour stimuli in area V2 of macaque monkeys. J Neurosci 24(13):3313–3324
3. Maunsell JH, van Essen DC (1983) Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. J Neurophysiol 49(5):1127–1147
4. Scholl B, Burge J, Priebe NJ (2013) Binocular integration and disparity selectivity in mouse primary visual cortex. J Neurophysiol 109(12):3013–3024
5. Ziemba CM, Freeman J, Movshon JA, Simoncelli EP (2016) Selectivity and tolerance for visual texture in macaque V2. Proc Natl Acad Sci 113(22):E3140–E3149
6. Komatsu H, Ideura Y, Kaji S, Yamane S (1992) Color selectivity of neurons in the inferior temporal cortex of the awake macaque monkey. J Neurosci 12(2):408–424
7. Marčelja S (1980) Mathematical description of the responses of simple cortical cells. J Opt Soc Am 70(11):1297–1300
8. Ohzawa I, DeAngelis GC, Freeman RD (1990) Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. Science 249(4972):1037–1041
9. Simoncelli E, Heeger D (1998) A model of neuronal responses in visual area MT. Vis Res 38(5):743–761
10. Lucas BD, Kanade T (1981) An iterative image registration technique with an application to stereo vision. In: Proceedings of Imaging Understanding Workshop, pp 121–130
11. Fraser A, Wilcox KJ (1979) Perception of illusory movement. Nature 281(5732):565–566
12. Hsieh PJ, Caplovitz GP, Tse PU (2006) Illusory motion induced by the offset of stationary luminance-defined gradients. Vis Res 46:970–978
13. Hayashi Y, Ishii S, Urakubo H (2014) A Computational model of afterimage rotation in the peripheral drift illusion based on retinal ON/OFF responses. PLoS ONE 9(12):e115464

14. Lindeberg T (1998) A scale selection principle for estimating image deformations. Image Vis Comput 16(14):961–977

15. Young RA, Lesporance RM, Meyer WW (2001) The Gaussian derivative model for spatial-temporal vision: I. Cortical model. Spat Vis 14:261–319

16. Lindeberg T (2013) A computational theory of visual receptive fields. Biol Cybern 107(6):589–635

17. Lindeberg T (2016) Time-causal and time-recursive spatio-temporal receptive fields. J Math Imaging Vis 55(1):50–88

18. Carandini M, Heeger D (2011) Normalization as a canonical neural computation. Nat Rev Neurosci 13(1):51–62

19. Boyraz P, Treue S (2011) Misperceptions of speed are accounted for by the responses of neurons in macaque cortical area MT. J Neurophysiol 105(3):1199–1211

20. Krekelberg B, van Wezel RJA, Albright TD (2006) Interactions between speed and contrast tuning in the middle temporal area: implications for the neural code for speed. J Neurosci 26(35):8988–8998

21. Priebe NJ, Cassanello CR, Lisberger SG (2003) The Neural representation of speed in macaque area MT/V5. J Neurosci 23(13):5650–5661

22. Hunter JN, Born RT (2011) Stimulus-dependent modulation of suppressive influences in MT. J Neurosci 31(2):678–686

23. Bouguet J (2000) Pyramidal implementation of the Lucas Kanade feature tracker. Intel Corporation, Microprocessor Research Labs, Santa Clara