

Simple Speed Estimators Reproduce MT Responses and Identify Strength of Visual Illusion

Daiki Nakamura*, Shunji Satoh

Graduate School of Information Systems, The University of Electro-Communications, Tokyo,
182-8585 JAPAN
daiki@hi.is.uec.ac.jp, shunji@uec.ac.jp

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 based on the concept of velocity filtering (e.g. proposed by Simoncelli & Heeger, 1998). In this paper, we give a novel interpretation of the computational function of MT neurons; an MT neuron would 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 by computer simulation for 8^8 inputs were used to estimate human illusory perception. Psychophysical experiments show 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 input 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 and takes its maximum value when $X = X_0$, then many researchers tend to infer that the recording neuron would prefer to X_0 , which is designated as *preferred X*. From the viewpoint of signal processing, we might conclude that such neurons would be band-

* Corresponding author. daiki@hi.is.uec.ac.jp, +81-42-443-5649

1 pass filters against quantity X with its maximum gain at X_0 . Many computational mod-
2 els are based on the computational interpretation of *preferred X* or *X-filtering* [7–9].

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

7 The first objective of our research is to provide a novel interpretation of unimodal
8 functions of MT neurons, which respond strongly to particular velocity, v_0 , of moving
9 visual stimuli. A simple speed-estimator (a proposed MT neuron model in this article)
10 also shows unimodal properties; the estimator based on **the Lucas-Kanade method** [10]
11 is designed so that its output $\hat{v} = f(v)$ is as equal to the actual velocity v as possible,
12 like a radar gun, if $0 \leq v < v_0$, where v_0 is not the preferred speed but the upper limita-
13 tion for correct estimation. If a velocity of moving stimulus exceeds the limitation, as
14 $v_0 < v$, then the velocity estimator would fail to estimate the correct speed. Such ve-
15 locity estimators will show a unimodal property of $f(v)$ if output \hat{v} converges to zero
16 (no response) for overly fast v exceeding v_0 .

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

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

43 As described in this paper, (1) we give a novel interpretation of the computational
44 function of MT neurons; an MT neuron would be a simple speed estimator with an
45 upper limitation for correct estimation. Then we develop an MT neuron model **that is**

1 not based on brain science. Using it, we examine whether the model reproduces MT
2 responses of speed selectivity such as presented in Fig. 2, or not, (2) whether the model
3 explains the luminance dependence of drift illusion, or not, (3) and we obtain model
4 predictions for all possible patterns by numerical simulation. In addition, we compare
5 the model predictions to results obtained from psychological experimentation to eval-
6 uate the plausibility of our computational model.
7

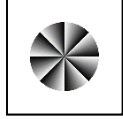

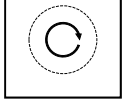
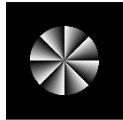


Prior stimuli	Post stimuli	Illusory motion
		
		

Fig. 1. Examples of drift illusion

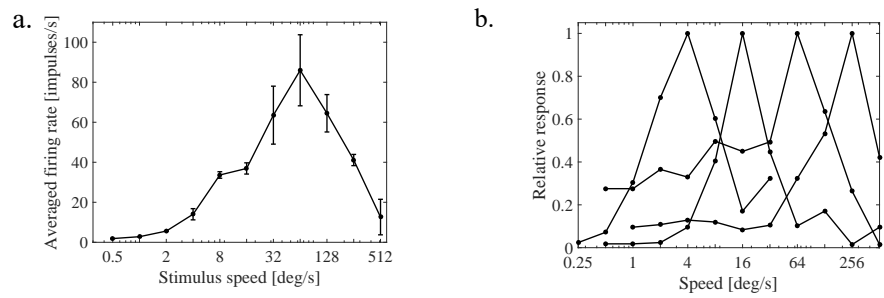


Fig. 2. Examples of MT neuron response [3]. (a) Solid lines show the firing rate of an MT neuron with respect to the speed of bar stimuli; bars represent the standard errors, (b) relative responses of four MT neurons at different speeds

1 2 Computational model of MT neurons

2 Simoncelli and Heeger proposed an MT neuron model based on the concept of *velocity*
 3 *filtering* [9]. An MT neuron of their model performs a spatio-temporal filter defined in
 4 the frequency domain.

5 As another interpretation of the computational role of MT neurons, we propose the
 6 following idea: an MT neuron would be a simple velocity estimator for which the output
 7 $\hat{\mathbf{v}} = (\hat{v}_x \ \hat{v}_y)^T$ (estimated velocity) is proportional to the true velocity \mathbf{v} of visual in-
 8 puts. We can derive such an estimator by minimizing an error function signified by
 9 $|\mathbf{v} - \hat{\mathbf{v}}|^2$, in which no concept of preferred velocity exists.

10 2.1 Basic Computation and MT model

11 We propose that the Lucas–Kanade (LK) method [10], a computer vision algorithm
 12 for optical-flow estimation minimizing a pre-defined error function, is the fundamental
 13 computation of MT cells. The LK method was derived under the following assump-
 14 tions:

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

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

$$21 \quad \hat{\mathbf{v}}(x, y, t) = (-1) \left\{ \begin{pmatrix} S_{xx}(x, y, t) & S_{xy}(x, y, t) \\ S_{xy}(x, y, t) & S_{yy}(x, y, t) \end{pmatrix} + \varepsilon^2 E \right\}^{-1} \begin{pmatrix} S_{xt}(x, y, t) \\ S_{yt}(x, y, t) \end{pmatrix} \quad (1)$$

$$22 \quad S_{ij}(x, y, t) \stackrel{\text{def}}{=} w(x, y) * \left\{ \frac{\partial I(x, y, t)}{\partial i} \frac{\partial I(x, y, t)}{\partial j} \right\} \quad (i, j = x, y, \text{ or } t) \quad (2)$$

23 In that equation, $I(x, y, t)$ represents the relative luminance of the input image in the
 24 (x, y) spatial coordinate system at time t , E is an identity matrix, $*$ is the convolution
 25 operator, and $\varepsilon^2 = 1.0 \times 10^{-4}$ is scalar parameter to avoid division by zero. Note that
 26 there is another implementation to avoid division by zero [14]. The $w(x, y)$ is a Gauss-
 27 ian window with standard deviation $\sigma_w = 11/6$ (window size is 11×11 pixels). The
 28 partial derivatives $\partial/\partial x$ and $\partial/\partial y$ for directional derivative are realized by numerical
 29 convolution between image I and the Gaussian derivative kernels of the spatial domain
 30 with size of $k \times k$ pixels [15-17]. The standard deviation of Gaussian derivative, σ_d , is
 31 proportional to the kernel size k : $3\sigma_d = k/2$ pixel. The temporal derivatives $\partial/\partial t$ rep-
 32 resent the difference of two adjacent frames: $I(t) - I(t - 1)$. Speed estimation with
 33 various σ_d is equal to speed estimation with various spatial resolution. An estimator
 34 with a smaller σ_d (a smaller k) provides a spatially higher-resolution map of optical
 35 flows and it is suitable for small objects, but such an estimator cannot take an accurate
 36 estimation for fast movements beyond its upper limitation. In contrast, a larger σ_d (a

larger k) is effective for fast motion and for large objects at the sacrifice of spatial resolution. This tradeoff should be considered for accurate speed estimation.

Herein, we propose a novel 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 shown 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 written in the section 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 $\phi + 90^\circ$ components of flows.

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

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) are, respectively, the directional derivatives along the ϕ and $\phi + 90^\circ$ direction. Expanding eq. (3), \hat{v}_ξ is written as shown below.

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

Assume 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.

$$\text{MT}_\phi^{\text{norm}}(\xi, \eta, t) = \frac{\hat{v}_\xi(\xi, \eta, t)}{\max_{\nu} \hat{v}_\xi(\xi, \eta, t)} = \frac{1}{L} \hat{v}_\xi(\xi, \eta, t) \quad (5)$$

Therein, $L = \max_{\nu} \hat{v}_\xi(\xi, \eta, t)$ is introduced for normalization of neural activities [18].

We determine the constant value of L using moving random dots. Hereafter, 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.

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

5 2.2 Numerical simulation

6 Using moving random dots, we simulated speed estimation by eq. (4) with respect to
 7 stimulus speed to examine whether the estimated speed presents a unimodal profile. An
 8 input image composes of Gaussian random dots of which each pixel value is drawn
 9 from the standard normal distribution. The image size was 150×150 pixels. Then we
 10 prepared 20 sets of input images for each speed. The motion was limited in the x-axis
 11 direction (zero degree; rightward motion). Hereinafter, we show the temporal average
 12 of $\hat{v}_x(0, 0, t)$, which is assumed to be proportional to the firing rate of an MT neuron
 13 with a receptive field on the center of images. The Gaussian derivative kernel size $k \times$
 14 k was 5×5 pixels ($3\sigma_d = 5/2$).

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

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

26 2.3 Kernel size and MT profile

27 Figure 2(b) shows that different MT neurons possess different peak speeds. We show
 28 that such response curves emerge from setting different kernel size k . Fig. 4 portrays
 29 response curves for kernel sizes of four kinds based on the octave concept: $k_i = 2^{i+1} +$
 30 $1, i = 1, 2, \dots, N$. In this article, we set $N = 4$ and $k_1 = 5, k_2 = 9, k_3 = 17, k_4 = 33$
 31 pixel. The left panel of Fig. 4 (linear plot) is averaged as $\hat{v}_x(k_i)$, whereas the right panel
 32 (semi-log plot) is averaged $MT_0^{\text{norm}}(k_i)$. The normalizing factor L_i in $MT_0^{\text{norm}}(k_i)$ is
 33 $L_i = \max_v \hat{v}_x(k_i)$, for example, $L_1 \approx 1.1, L_2 \approx 1.5$, as shown on the left panel of Fig.
 34 4.

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

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

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

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

8 Therefore, we propose the following computational interpretation of the MT popula-
9 tion: the vision system employs numerous MT neurons with different properties be-
10 cause of the tradeoff between spatial resolution and speed limitation.

11 2.4 Read out from MT population

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

$$18 \quad \bar{v} = \frac{1}{N} \sum_{i=1}^N \hat{v}(k_i) \quad (6)$$

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

$$21 \quad \bar{v} = \frac{1}{N} \sum_{i=1}^N L_i \cdot \text{MT}^{\text{norm}}(k_i) \quad (7)$$

22 The model of eq. (7) is coincidentally identical to that proposed by Boyraz and Treue
23 [19], whose model accounts for the input-size effect on perceptual speeds. In section
24 5.1, we discuss the relation of eq. (7), Boyraz and Treue model, and other read-out
25 models.

26 2.5 Discussion

27 MT neurons have been believed to tune for their preferred speed. However, response
28 curves of the MT model based on the LK method, which is a simple speed estimator,
29 also presents unimodal functions similar to MT response curves. The MT model could
30 not correctly estimate stimulus speed because exceeding a specific speed is a violation
31 of assumption (b) of the LK method: “a change in luminance can be expressed by first-
32 order approximation of the Taylor expansion.” Therefore, we can rephrase the state-
33 ment of *preferred speed* by *upper limitation* of correct estimation.

34 Our examination revealed that the speed estimator based on the LK method also
35 reproduced that MT neurons reached its maximum firing rate at various speeds, as
36 shown in Fig. 4, using various kernel sizes for calculating the spatial derivative. This
37 result demonstrates that each MT neuron estimates an optical flow with various kernel

1 sizes. It is possible for the normalized LK model (eq. (5)) to be constructed as a neural
2 network using V1 neuron models that calculate spatio-temporal derivative [15-17].

3 We successfully reproduce the unimodal profile of MT outputs with respect to input
4 stimulus. However, we recognized that the current model is not sufficient to explain for
5 complex properties of MT neurons, e.g., contrast dependency [20], spatial frequency
6 dependency [21], and texture dependency [22]. Those problems are included as future
7 works.

8
9

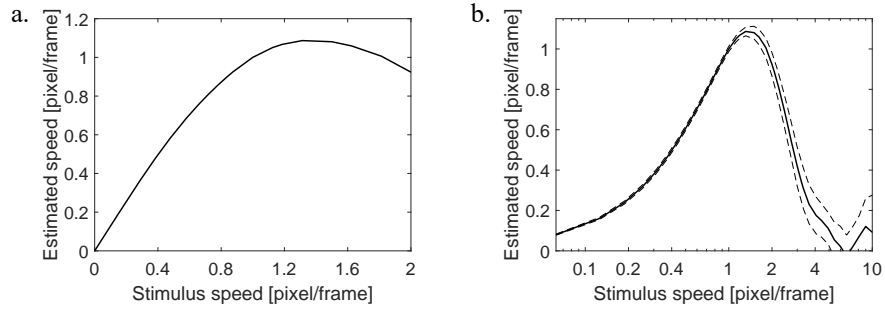


Fig. 3. Averaged speed of model-estimated speed \hat{v}_x at different stimulus speeds: (a) the horizontal axis is linear and (b) the horizontal axis is logarithmic

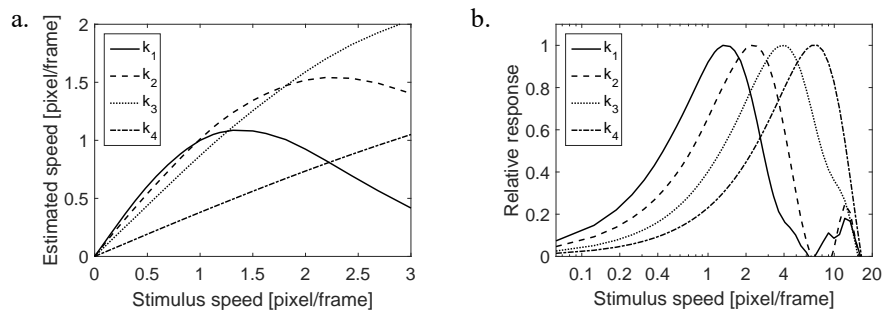


Fig. 4. Averaged estimated speeds obtained using various kernel sizes for calculating **spatio-temporal** derivative: (a) the horizontal axis is linear and the (b) relative response normalized to its maximum value as logarithmic scales for the horizontal axis

3 Reproducing rotational illusion dependent on background luminance

We examine the plausibility of our read-out model (eq. (6) or eq. (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) would be spatially rotating and that the direction of rotation would depend 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 gray scale images of which luminance values were 0.0 (darkest, black) to 1.0 (brightest, white). Figure 7 presents the estimated optical flow vectors obtained using eq. (6) ($k_i \in \{5, 9, 17, 33\}$). In Fig. 7, clockwise rotation vectors appeared when the relative background luminance was 1.0 (Fig. 7, top). In contrast, counterclockwise rotation vectors appeared when the relative background luminance of 0.0 (Fig. 7, 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, as known as the rot operator introduced into vector analysis.

$$\bar{R} = \frac{1}{|S|} \iint_S \text{rot}_{2D} \bar{\mathbf{v}}(x, y, t) dS = \frac{1}{|S|} \iint_S \frac{\partial \bar{v}_y(x, y, t)}{\partial x} - \frac{\partial \bar{v}_x(x, y, t)}{\partial y} dS \quad (8)$$

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. Fig. 8 shows 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 positive largest value for counterclockwise rotation was obtained at minimum relative luminance ($I = 0.0$). The magnitude of rotation was zero at background luminance $I = 0.5$.

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.

1 Let us consider the theoretical reason for luminance dependence of illusory rotation.
 2 From eq. (1), we ascertained that the temporal derivative term $\partial I/\partial t$ affects the rota-
 3 tional direction and rotational strength. For simplicity, we analyzed the illusion on the
 4 polar coordinate system (r, θ) using the center of FW stimuli as the origin (Fig. 9a).
 5 The right panel of Fig. 9 presents the relative luminance $I(r, \theta)$ of FW stimuli with
 6 respect to angle θ . The direction of optical flows is almost an angular direction. We
 7 restrict ourselves to consider the case of $\theta = 0 \sim 45^\circ$ because FW stimuli are composed
 8 of periodic circular sectors of 45 deg. In the case of left panel of Fig. 9 ($\phi = \theta$), eq. (3)
 9 is rewritten as

$$10 \quad \begin{pmatrix} \hat{v}_\xi(r, \theta, t) \\ \hat{v}_\eta(r, \theta, t) \end{pmatrix} = (-1) \begin{pmatrix} 0 \\ \frac{\partial I(r, \theta, t)}{\partial t} / \left\{ \frac{1}{r} \frac{\partial I(r, \theta, t)}{\partial \theta} \right\} \end{pmatrix}. \quad (9)$$

11 Herein, the luminance change of radial direction is zero ($\partial I/\partial \xi = 0$), the window is
 12 the Dirac delta function ($w(r, \theta) = \delta(r, \theta)$). The parameter of avoiding zero division
 13 is zero ($\varepsilon^2 = 0$). From eq. (10), the estimated angular velocity $\omega(r, \theta, t)$ is calculated
 14 using the following formula

$$15 \quad \omega(r, \theta, t) = \frac{1}{r} \hat{v}_\eta(r, \theta, t) = - \frac{\partial I(r, \theta, t)}{\partial t} / \frac{\partial I(r, \theta, t)}{\partial \theta}. \quad (10)$$

16 Eq. (11) shows that the sign of temporal change of luminance (numerator) and the sign
 17 of spatial change (denominator) determine the direction of rotation. Figure 10 portrays
 18 $\omega(\theta, t)$, $\partial I(\theta, t)/\partial t$, and $\partial I(\theta, t)/\partial \theta$ under background luminance of 0.0 (black) and
 19 1.0 (white). Comparing the two columns in Fig. 10, the temporal derivative term causes
 20 rotational direction and rotational strength of drift illusion's dependency on background
 21 luminance. Only the temporal derivative term is dependent on background luminance.

22 The success of those analyses is attributable to apply the simple formula composed of
 23 spatio-temporal derivatives to the fundamental computation of MT cells.

24 Hsieh et al. concluded that illusory motion might be related to the afterimage by psy-
 25 chophysical experiments using visual inputs similar with FW stimuli (Hsieh et al.
 26 2006). As an alternative explanation of visual illusion for the FW stimuli, we showed
 27 that the illusion might be caused by incorrect estimation of optical flows by MT neurons
 28 (Fig. 7). In this simulation, illusory optical flows that related to illusory rotation ap-
 29 peared on just one frame because the temporal derivatives $\partial/\partial t$ were implemented by
 30 the difference of two adjacent frames: $I(t) - I(t - 1)$. That is, illusory motion appears
 31 just after disappearing visual stimuli, and the duration of illusory motion by the model
 32 is the frame interval. Actually, Hsieh et al. concluded that motion illusion lasts shorter
 33 than the decay rate of afterimage, and that the illusion observed only at the beginning
 34 phase of disappearing the FW-type stimuli. When $\partial/\partial t$ of the LK model is imple-
 35 mented by the temporal convolution kernel formulated by the Gaussian derivative with
 36 standard deviation σ_t (Young et al., 2001, Lindeberg も入れる?), the duration of illu-
 37 sory optical flows is proportional to σ_t . Consequently, the model properties are con-
 38 sistent with the conjectures by Hsieh et al.

39

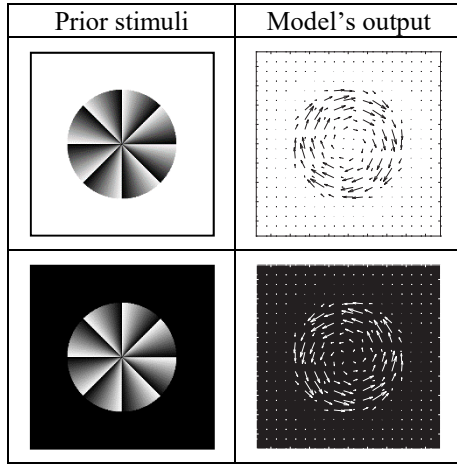


Fig. 5. Output vectors (optical flow, estimated perception of motion) obtained from our read-out model (eq. (6))

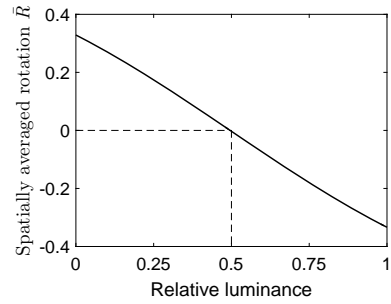


Fig. 6. Rotations of model outputs with respect to the relative luminance

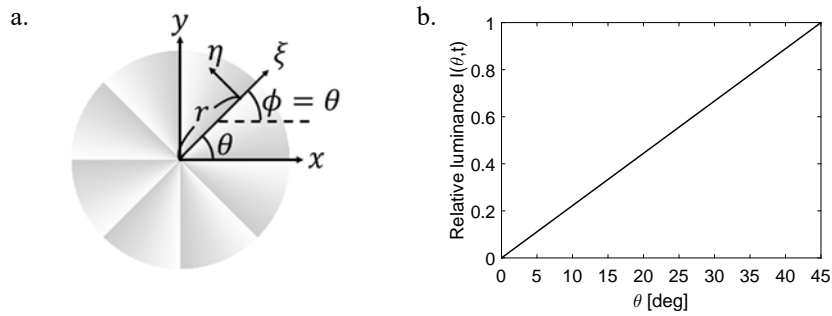


Fig. 7. Relative luminance of FW stimuli on the polar coordinate system $I(\theta)$. (a) FW stimulus and ξ, η axis, (b) relative luminance of FW stimulus with respect to the polar angle $I(\theta)$

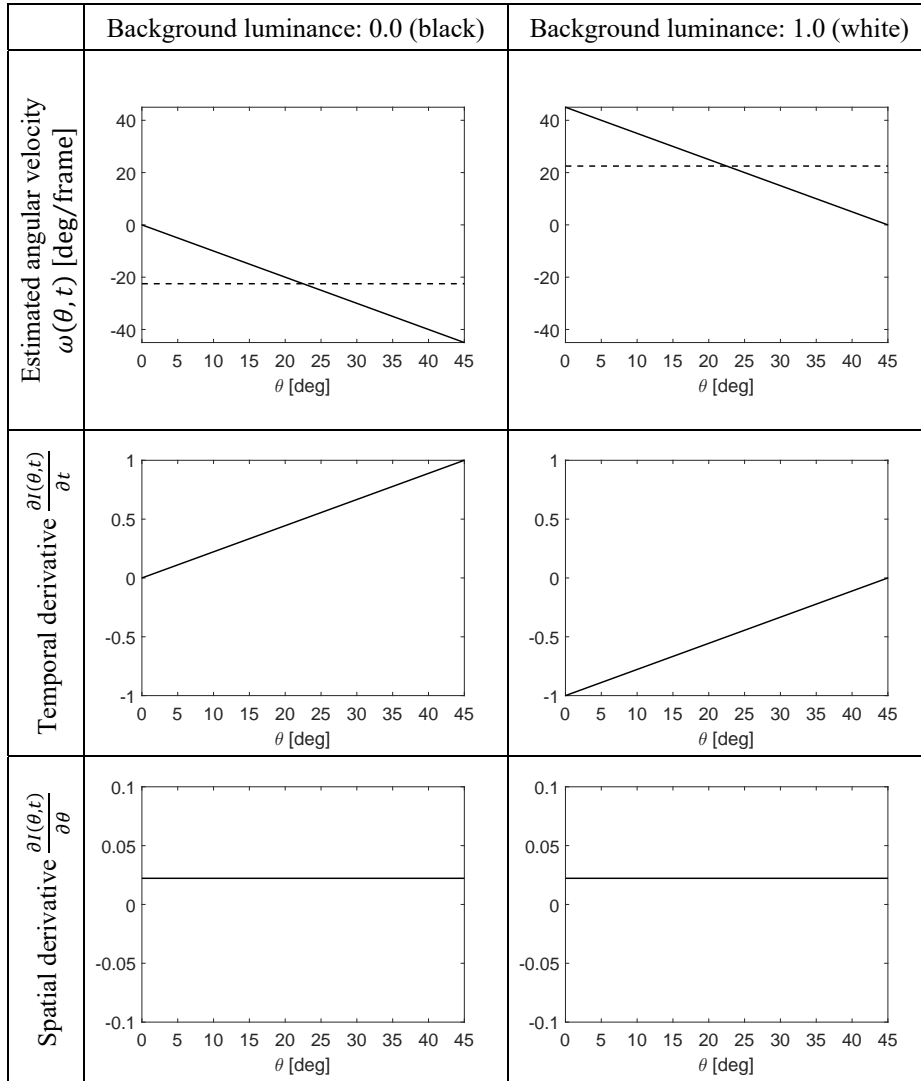


Fig. 8. Cause of drift illusion dependence on background luminance

1 **4 Model predictions and psychological experiments**

2 We next evaluate the correlation between human perception and model prediction using prior/post images with white background to examine the generalization ability of the MT model.

5 **4.1 Circular stimulus**

6 As portrayed in Fig. 12, a prior stimulus comprises circular sectors of 45° . A luminance pattern in a circular sector comprises eight gray levels: a combination of $I \in \{0/7, 1/7, 2/7, \dots, 7/7\}$. The number of possible patterns is $8^8 = 16,777,216$.

9 **4.2 Selection of stimuli for psychological experiment**

10 We obtain the rotation \bar{R} for over 16 million kinds of stimuli of a white background. To reduce simulation time, we set $N = 1$ and $k = 5$ pixel in read-out model. We will discuss model predictions using other kernels in section 4.5. Fig. 11 shows a histogram of rotation \bar{R} , indicating that almost all stimuli have small rotation $|\bar{R}| \approx 0$, 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 hr (dual processor Xeon E5-2630 v2 2.6 GHz, Intel Corp.).

18 For psychological experiments, 33 patterns were chosen randomly from 16 million patterns so that the model predictions \bar{R} were distributed uniformly, and that a selected pattern contains both black and white ($I = 0.0$ and $I = 1.0$). Real values of Fig. 12 signify \bar{R} s from -0.0100 to 0.0136 .

22 **4.3 Methods**

23 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 $22.0 \times 16.6^\circ$. The circular stimulus diameter was 13.0° (300 pixels). The maximum luminance (white; $I = 1.0$) was 81.3 cd/m^2 .

29 The 33 prior stimuli in Fig. 12 were displayed randomly. Each stimulus was displayed 10 times. Post stimuli were uniformly white. Prior stimuli were presented for 1500 ms. Subsequently, prior stimuli disappeared and 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 ethical committee of the University of Electro-Communications.

1 4.4 Correlation between model output and human perception

2 The fractional number in Fig. 12 is the probability of human judgment for “clockwise”
3 rotation for 50 trials. For example, 49/50 of #1 means that humans tend to perceive
4 clockwise illusory rotation, and 2/50 of #33 perceive counterclockwise rotating illu-
5 sion.

6 Next, we compared model predictions with human responses. We adopt the following
7 formula to transform rotation \bar{R} into the stochastic judgment of clockwise motion
8 $\Pr(CW)$.

$$9 \quad \Pr(CW) = \frac{1}{2} \left(1 - \operatorname{erf} \left(\frac{\bar{R}}{s\sqrt{2}} \right) \right) \quad (11)$$

10 Therein, $\operatorname{erf}()$ is the error function; s is a positive parameter. We assumed that the
11 chance level corresponds to circumstances in which $\bar{R} = 0$ and $\Pr(CW) = 0.5$. The
12 free parameter s of eq. (12) was determined by application of a nonlinear fitting of the
13 model function $\Pr(CW)$ to 33 data of human judgment. The best parameter was $s =$
14 0.013 .

15 Figure 13 presents a scatter plot of model judgment and human judgment. The open
16 circle corresponds to results for FW stimuli drawn with eight grayscale levels. Real
17 values r and p in the upper left of Fig. 13, are Pearson’s correlation coefficient and p
18 value for testing the hypothesis of no correlation. If the model prediction were perfectly
19 correct, then markers in Fig. 13 would be arranged on the diagonal line. The computa-
20 tional prediction of human perception was not perfect, but a positive correlation be-
21 tween them might be readily apparent (0.81 of correlation coefficient and $p < 10^{-8}$
22 for no correlation testing). We obtained illusory patterns aside from the FW pattern, as
23 shown in #1 and #33 of Fig. 11.

24 4.5 Effects of kernel size on model predictions

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

31 Solid lines of Fig. 14 show correlation coefficients r between the model judgment and
32 human judgment with respect to kernel size k . Dashed lines are the correlation coeffi-
33 cient in the case of $N = 4$. Input images were scaled to obtain different image sizes
34 using scale factors $f \in \{1/4, 1/2, 1/1\}$. From Fig. 14, it is apparent that larger kernel
35 size is better for larger input image. When $f = 1/2$, $N = 1$, and $k = 17$, the best cor-
36 relation coefficient of 0.96 is obtained.

37 Figure 15 presents a scatter plot of model judgment and human judgment using the
38 best parameters. Comparison of Fig. 15 and Fig. 13 shows improvement of the r value.

1 These results indicate that a kernel size selection according to image size is an important
2 computation accounting for visual perception.
3 Using multiple kernel size and the read-out by eq. (6) scores a better r value, on aver-
4 age. The average r of multiple kernels and single kernel were, respectively, 0.913 and
5 0.864. Object sizes and the best kernel sizes are factors that are unknown in advance.
6 Therefore, a model using multiple kernels is expected to be useful in general cases to
7 achieve rapid estimation.
8
9

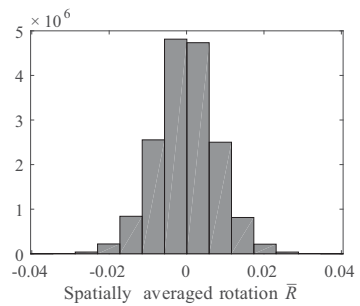


Fig. 9. Histogram of spatially averaged rotation \bar{R} for 16,777,216 stimuli













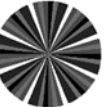






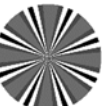














Ex -0.0189 50/50 		#1 -0.0079 49/50 	#2 -0.0036 49/50 	#3 -0.0075 49/50 	#4 -0.0044 48/50 	#5 -0.0081 47/50 
#6 -0.0100 45/50 	#7 -0.0096 44/50 	#8 -0.0035 37/50 	#9 -0.0019 36/50 	#10 -0.0033 35/50 	#11 -0.0015 33/50 	#12 -0.0035 32/50 
#13 -0.0020 31/50 	#14 -0.0015 26/50 	#15 0.0029 23/50 	#16 0.0002 23/50 	#17 0.0006 20/50 	#18 -0.0035 18/50 	#19 0.0006 17/50 
#20 0.0001 16/50 	#21 0.0046 16/50 	#22 0.0050 13/50 	#23 -0.0069 13/50 	#24 0.0070 8/50 	#25 0.0136 6/50 	#26 0.0030 6/50 
#27 0.0038 5/50 	#28 0.0071 5/50 	#29 0.0083 4/50 	#30 0.0113 3/50 	#31 -0.0005 2/50 	#32 -0.0003 2/50 	#33 0.0054 2/50 

Fig. 10. Stimuli used in psychological experiments sorted by the probability of human judgment. #1–#33 are the indexes of stimuli, and Ex. is FW stimulus drawn with eight grayscale level. Negative and positive real values are spatially averaged rotation \bar{R} . Fractional numbers are the probability of human judgment to clockwise rotation of perception for 50 trials.

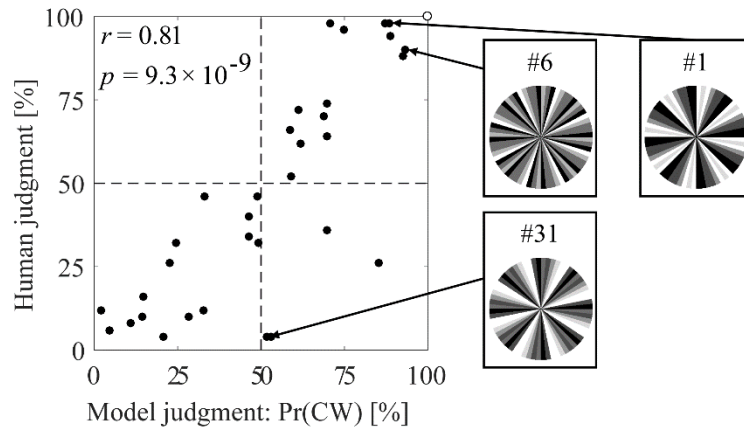


Fig. 11. Scatter plot of model judgment and human judgment with $N = 1$ and $k = 5$. An open circle at the upper right corner of the plot corresponds to the original FW stimulus drawn with eight grayscale levels, r is Pearson's correlation coefficient, and p is the p value for testing the no-correlation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

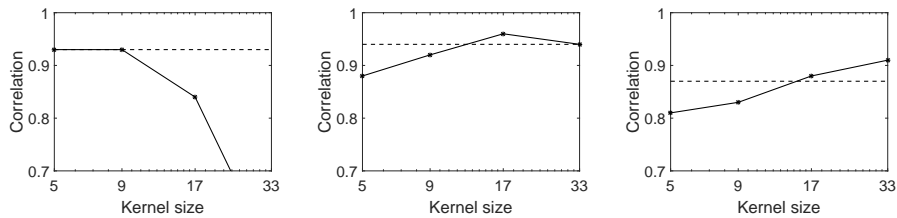


Fig. 12. Correlation coefficients between human judgment and model judgment using single (solid line) and multiple (dashed line) kernels with resolution factor $f = 1/4$ (left), $f = 1/2$ (center), $f = 1/1$ (right; original scale).

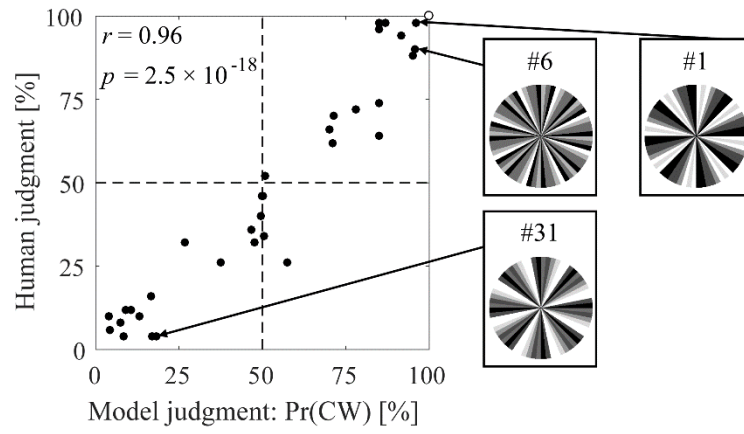


Fig. 13. Scatter plot of model judgment and human judgment of the best parameter ($N = 1$, $k = 17$, and $f = 1/2$)

1 5 General discussion and conclusion

2 We demonstrated that the response curves of the MT model based on the Lucas-
3 Kanade method also show unimodal functions with respect to stimulus speed such as
4 MT neuron response curves, although the model was not formulated to show unimodal
5 responses. Our read-out model from MT population accounted for human illusory per-
6 ception. First, in this section, we evaluate the model by comparison to the other char-
7 acteristics of physiology and by comparison to other computational model of MT neu-
8 rons. Second, we present clues to discover the novel illusory patterns.

9 MT model and read-out model

10 The tuning width, which is a full width at half maximum of tuning curve, is another
11 aspect to evaluate the plausibility of the MT model. Maunsell and van Essen reported
12 that the average tuning width for speed of MT neurons is approximately a 7.7-fold
13 change of speed (2.9 octaves) [3]. The tuning width of the normalized LK model
14 $MT_0^{\text{norm}}(k = 5)$ (the smallest k ; Fig. 3) is a 6.4 fold change of speed (2.7 octaves). The
15 tuning width similarity is expected to support the plausibility of the LK model.

16 Assuming that MT neurons are velocity estimators, we obtained another interpretation
17 of the peak speed of the MT tuning curve. It does not mean a *preferred* speed but an
18 *upper limitation* for correct estimation of speed. Herein, we try to present a computa-
19 tional explanation of complex responses of MTs depending on the stimulus properties.
20 Krekelberg et al. reported that the peak speed of MT neurons decreased with lower
21 contrast of displayed stimulus [21]. This phenomenon is expected to be trivial because
22 a lower-contrast input causes a lower signal-to-noise (SN) ratio. Consequently, the up-
23 per limitation for correct estimation also decreases for signals with a lower SN ratio.
24 The side effect of parameter ε^2 of eq. (1) is also related to the contrast dependence of
25 peak speeds.

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

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

$$35 \quad v_p = \frac{\sum_{i=1}^N (MT_i^{\text{norm}} \times L_i)}{\alpha} \quad (12)$$

36 Herein, v_p stands for the perceived speed (result of read-out from MT population), N
37 signifies the number of MT neurons, MT_i^{norm} denotes the relative response of an MT
38 neuron, L_i represents a specific value (usually designated as “label”) with a specific
39 MT neuron, and α is a normalizing factor. Changing normalizing factor α in eq. (13),
40 we can express the three models: $\alpha = N$ corresponds to our read-out model, $\alpha = \text{const.}$
41 corresponds to Boyraz and Treue model. The original vector averaging model is given
42 as $\alpha = \sum_{i=1}^n MT_i^{\text{norm}}$. Boyraz and Treue did not describe the computational meaning of

1 introducing a constant α . Their model (constant α) reproduced misperceptions of speed
2 perception dependent on the stimulus size. It is noteworthy that the model of constant
3 α by Boyraz and Treue is computationally equivalent to our simple read-out model of
4 eq. (7), averaging estimated speeds, in which $\alpha = N$ is also a constant value. Therefore,
5 the size dependence of motion perception can also be interpreted as a side effect of our
6 read-out model.

7 5.1 Illusory motion perception

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

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

25 5.2 Conclusions

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

35 Results of this study can elucidate visual systems from various aspects, facilitate the
36 evaluation of various vision models, and help to generate new illusory patterns.

37 There are several other variations of the LK method [23]. MT models based on the
38 other methods might also reproduce MT responses and human motion perceptions. It is
39 future work to distinguish which algorithm is the most suitable for the MT model.

Acknowledgments

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. *The Journal of Physiology* 160:106–154
2. Ito M, Komatsu H (2004) Representation of angles embedded within contour stimuli in area V2 of macaque monkeys. *The Journal of Neuroscience* 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. *Journal of Neurophysiology* 49(5):1127–1147
4. Scholl B, Burge J, Priebe NJ (2013) Binocular integration and disparity selectivity in mouse primary visual cortex. *Journal of Neurophysiology* 109(12):3013–3024
5. Ziemba CM, Freeman J, Movshon JA, Simoncelli EP (2016) Selectivity and tolerance for visual texture in macaque V2. *Proceedings of the National Academy of Sciences* 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. *The Journal of Neuroscience* 12(2):408–424
7. Marčelja S (1980) Mathematical description of the responses of simple cortical cells*. *Journal of the Optical Society of America* 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. *Vision Research* 38(5):743–761
10. Lucas BD, Kanade T (1981) An Iterative Image Registration Technique with an Application to Stereo Vision. *Proceedings of Imaging Understanding Workshop*: 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. *Vision Research* 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 and Vision Computing* 16(14): 961–77.
15. Young RA, Leporance RM, Meyer WW (2001) The Gaussian derivative model for spatial-temporal vision: I. Cortical model. *Spatial Vision* 14:261–319
16. Lindeberg T (2013) A Computational Theory of Visual Receptive Fields. *Biological Cybernetics* 107(6): 589–635.
17. Lindeberg T (2016) Time-Causal and Time-Recursive Spatio-Temporal Receptive Fields.” *Journal of Mathematical Imaging and Vision* 55 (1): 50–88.

18. Carandini M, Heeger D. (2011) Normalization as a Canonical Neural Computation. *Nature Reviews Neuroscience* 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. *The Journal of Neuroscience*: 26(35):8988–8998
21. Priebe NJ, Cassanello CR, Lisberger SG (2003) The Neural Representation of Speed in Macaque Area MT/V5. *The Journal of Neuroscience* 23(13): 5650–61.
22. Hunter JN, Born RT (2011) Stimulus-Dependent Modulation of Suppressing Influences in MT. *The Journal of Neuroscience* 31(2): 678–86.
23. Bouguet J (2000) Pyramidal implementation of the Lucas Kanade feature tracker. Intel Corporation, Microprocessor Research Labs.