Motion illusions as optimal percepts

Yair Weiss[1] , Eero P. Simoncelli[2], and Edward H. Adelson[3]

[1] *School of Computer Science and Engineering, Hebrew University of Jerusalem, Israel 91904*

[2] *Howard Hughes Medical Institute, Center for Neural Science, and Courant Institute of Mathematical Sciences, New York University, New York, NY 10003*

[3] *Brain and Cognitive Sciences Department, Massachusetts Institute of Technology, 77 Massachussetts Ave, Cambridge, MA 02139*

The pattern of local image velocities on the retina encodes important environmental information. Psychophysical evidence reveals that while humans are generally able to extract this information, they can easily be deceived into seeing incorrect velocities. We show that these 'illusions' arise naturally in a system that attempts to estimate local image velocity. We formulate a model for visual motion perception using standard estimation theory, under the assumptions that (a) there is noise in the initial measurements, and (b) slower motions are more likely to occur than faster ones. A specific instantiation of such a velocity estimator accounts for a wide variety of psychophysical phenomena.

Introduction

The ability of humans to analyze visual motion in general scenes far exceeds the capabilities of the most sophisticated computer vision algorithms. Yet psychophysical experiments reveal that humans also make some puzzling mistakes, misjudging speed or direction of very simple stimuli. In this paper, we propose that such mistakes of human motion perception represent the best solution of a rational system designed to operate in the presence of uncertainty.

In both biological and artificial vision systems, motion analysis begins with local
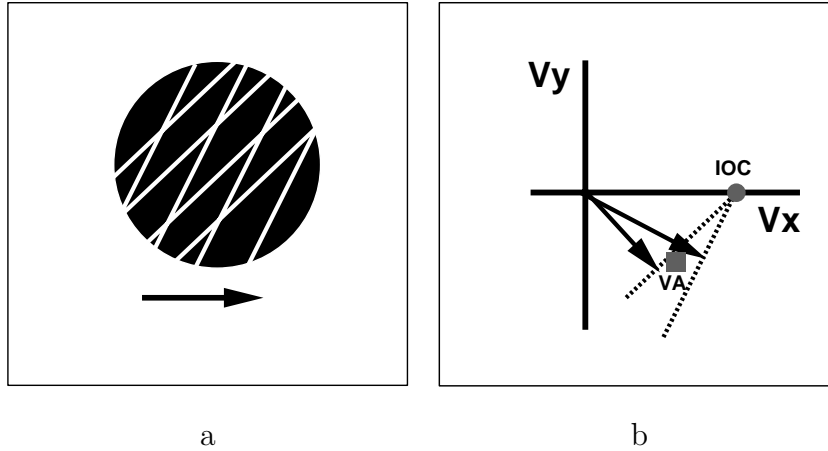
a   b

Fig. 1: Intersection of Constraints. **a.** Drifting gratings superimposed in the image plane produce a translating 'plaid' pattern. **b.** Due to the aperture problem, the measurements for a single grating are consistent with a family of motions all lying on a constraint line in the space of translational velocities. The Intersection of Constraints (IOC) solution is the unique velocity consistent with the constraint lines of both gratings. The Vector Averaging (VA) solution is the average of the two normal velocities. There is experimental evidence for both types of combination rule.

measurements (e.g., the output of direction selective cells in primary visual cortex[1], or spatial and temporal derivative operators[2, 3]). These are then integrated to generate larger, more global motion descriptions. The integration process is essential because the initial local motion measurements are ambiguous. For example, in the neighborhood of a contour, only the motion component perpendicular to the contour can be determined (a phenomenon referred to as the 'aperture problem')[5, 2, 6, 7]. Such an integration stage appears to be consistent with much of the psychophysical[8, 9, 10, 11] and physiological[8, 12, 13, 14] data.

Despite the vast amount of psychophysical data published over the past two decades, the nature of the integration scheme underlying human motion perception remains controversial. This is true even for the simple, widely studied 'plaid' stimuli, in which two superimposed oriented gratings translate in the image plane (Fig. 1a). Due to the aperture problem, each grating motion is consistent with an infinite number of possible translational velocities lying on a constraint line in the space of all velocities (Fig. 1b). When a single drifting grating is viewed in isolation, subjects typically perceive it as

translating in a direction normal to its contours (arrows in Fig. 1b). When two gratings are presented simultaneously, subjects often perceive them as a coherent pattern translating with a single motion[5, 7].

How is this coherent pattern motion estimated? Most explanations are based on one of three rules[7]: Intersection of Constraints (IOC), Vector Average (VA), or Feature Tracking (FT). The Intersection of Constraints (IOC) solution is the unique translation vector consistent with the information of both gratings. Graphically, this corresponds to the point in velocity space that lies at the intersection of both constraint lines (circle in Fig. 1b). The Vector Average (VA) solution is the average of the two normal velocities. Graphically this corresponds to the point in velocity space that lies halfway in between the two normal velocities (square in Fig. 1b). A Feature Tracking solution corresponds to the velocity of some feature of the plaid intensity pattern (for example, the locations of maximum luminance at the grating intersections)[15, 16]. For plaids, the FT and IOC solutions both correspond to the veridical pattern motion.

Which of the three rules best describes the perception of human observers? The answer is not clear: under different stimulus conditions, the perceived pattern motion can be nearly veridical (consistent with IOC or FT) or closer to the VA solution. The relevant factors include relative grating orientation and speed[17, 18, 19], contrast[20], presentation time[17], and retinal location[17].

Similar effects have been reported with stimuli that appear quite different from plaids[16, 21]. Consider a moving rhombus (Fig. 2). As with a plaid pattern, the motion of each opposing pair of sides is consistent with a constraint line in the space of velocities. As shown in the velocity space diagrams (Fig. 2c,f), IOC/FT predicts horizontal motion, while VA predicts diagonal motion. Perceptually, however, the rhombus appears to move horizontally at high contrast, but diagonally at low contrast. To further complicate the situation, the percept depends on the shape. If the rhombus is fattened (Fig. 2d), it appears to move horizontally at both contrasts. The reader may view these stimuli at http://neurosci.nature.com/web_specials.

In order to explain the appearance of these stimuli using the three rules, one is
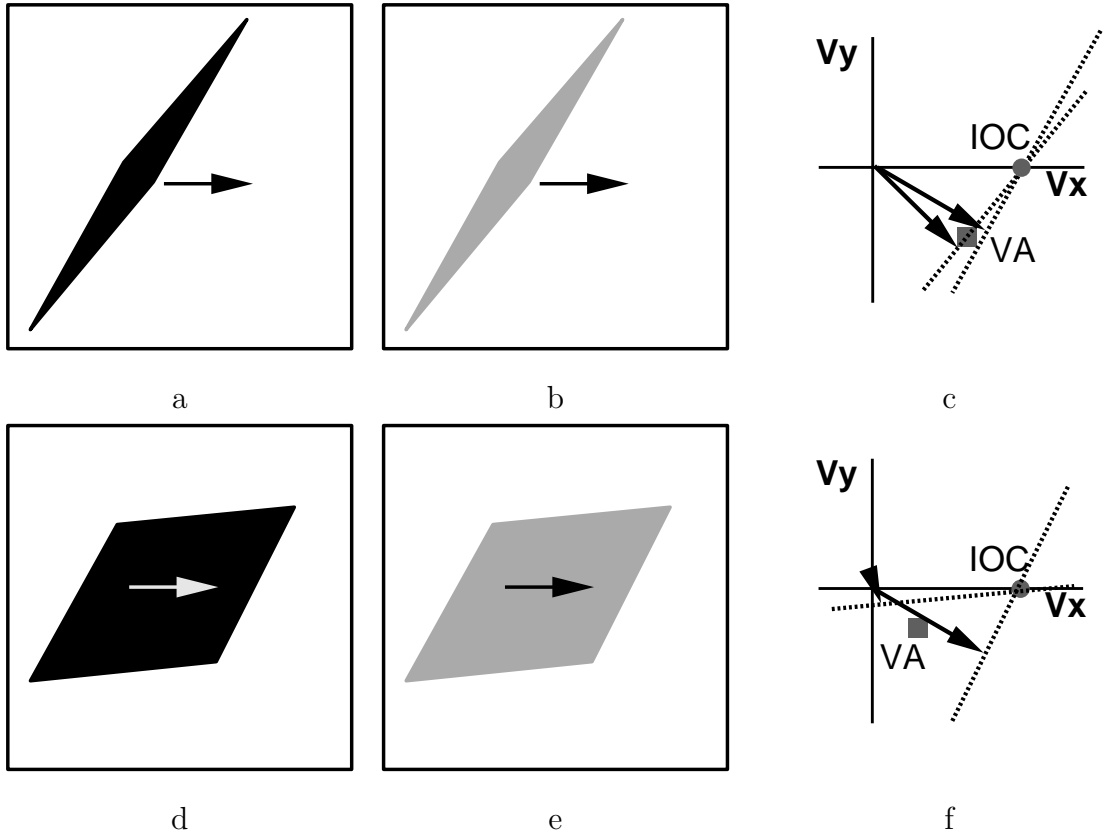
3

Fig. 2: Insufficiency of either VA, IOC or FT rules as an explanation for human perception of a horizontally moving rhombus. **a.** A 'narrow' rhombus at high contrast appears to move horizontally (consistent with IOC/FT). **b.** A narrow rhombus at low contrast appears to move diagonally (consistent with VA). **c.** Velocity space constraints for a narrow rhombus. **d-e.** A 'fat' rhombus at low or high contrast appears to move horizontally (consistent with IOC/FT). **f.** Velocity space constraints for a fat rhombus.

tempted to propose that the visual system uses VA for the thin low-contrast rhombus, and IOC/FT for the thin high-contrast rhombus and the fat rhombus. Although a model based on such an ad-hoc combination of rules can certainly fit the data, it is clearly not a parsimonious explanation. Furthermore, each of the idealized rules is limited to stimuli containing straight structures at only two orientations, and does not prescribe a method for computing the normal velocities of those structures. One would prefer a single, coherent model that can predict the perceived velocity of any arbitrary spatiotemporal stimulus that is seen as translational. In the following section, we develop a such a model based on a simple formulation of the problem of velocity estimation and a few reasonable assumptions.

Bayesian Model

In Helmholtz's view, our percepts are our best guess as to what is in the world, given both our sensory data and our prior experience[22]. In order to make this definition more quantitative, one must specify (a) what is 'best' about a best guess, and (b) the way in which prior experience should influence that guess. In the engineering literature, the theory of estimation provides a formalization of these concepts. The simplest and most widely known estimation framework is based on Bayes' rule (see[23] for examples of Bayesian models in perception and[24, 25] for Bayesian motion models). Following an approach described in previous work [26, 27, 28, 29], we develop an optimal Bayesian estimator (known as an 'ideal observer' in the psychophysics literature) for two-dimensional velocity. For the purposes of this paper, we limit ourselves to cases for which humans see a single global translational motion (i.e, no deformation, rotation, occlusion boundaries, transparency, etc). The majority of psychophysical experiments on the aperture problem focus on such cases, as do the majority of models (although we have developed extensions of the model described here that can handle more complicated scenes[29]).

Our model begins with the standard principle of intensity conservation: we assume that any changes in image intensity over time are due entirely to translational motion

of the intensity pattern. We then make two basic assumptions: (1) local image measurements are noisy and (2) image velocities tend to be slow. Below, we formulate these assumptions using probability distributions, and used Bayes' rule to derive the ideal observer. Mathematical details are provided in **Methods**.

The first assumption is instantiated using a noise model that is widely used in engineering because of the tractability of the solution: measurements are contaminated with additive, independent Gaussian noise, with a known standard deviation, σ. Although this simple noise model is unlikely to be correct in detail, we show in the next section that it is sufficient to account for much of the data. The noise model provides a functional form for the local likelihood: a distribution over the space of velocities based on measurements made in a local image patch. Figure 3 depicts this likelihood as a gray-level image, in which intensity corresponds to probability. For patches containing a single edge, the likelihood function is similar to a 'fuzzy' constraint line - velocities on the constraint line have highest likelihood, and the likelihood decreases with distance from the line. The 'fuzziness' of the constraint line is governed by σ, the standard deviation of the assumed noise. At corners, where local motion measurements are less ambiguous, the likelihood no longer has the elongated shape of a constraint line but becomes tightly clustered around the veridical velocity.

The model of additive Gaussian noise also leads to an interesting dependence of the likelihood on contrast. For a fixed noise level, σ, the likelihoods are broader at low contrast (bottom of Fig. 3). This makes intuitive sense: at low contrast there is less information regarding the exact speed of the stimulus, so there is more local uncertainty and the likelihood is more spread out. In the extreme case of zero contrast, the uncertainty is infinite.

The second assumption underlying our ideal observer is that velocities tend to be slow. Suggestions that human observers prefer the 'shortest path' or slowest motion consistent with the visual input date back to the beginning of the 20th century (see Ref. [30] and references within). In particular, Wallach (1935) hypothesized that humans prefer to see the normal velocity for a single line segment because that is the slowest velocity
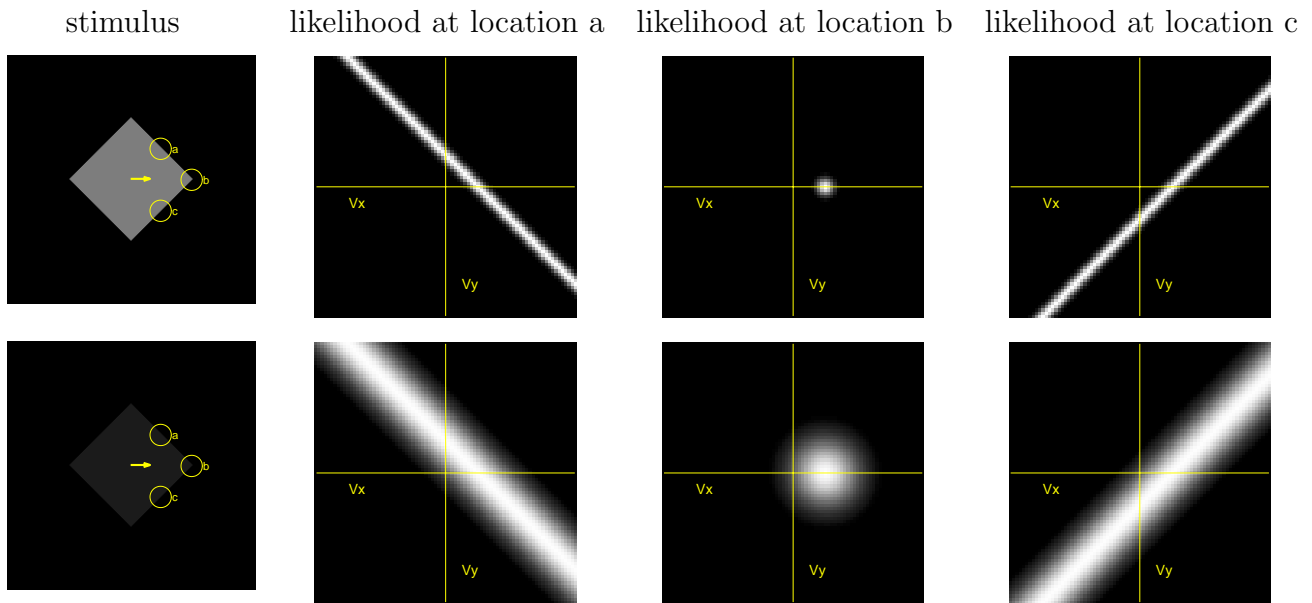
stimulus	likelihood at location a	likelihood at location b	likelihood at location c

Fig. 3: Likelihood functions for three local patches of a horizontally translating diamond stimulus, computed using Eq. (4). Intensity corresponds to probability. **Top:** high contrast sequence. **Bottom:** Low contrast sequence, with the same parameter σ. At an edge the local likelihood is a 'fuzzy' constraint line, while at corners the local likelihood is peaked around the veridical velocity. The sharpness of the likelihood decreases with decreasing contrast.

consistent with the image data[5]. Likewise in apparent motion displays humans tend to choose the shortest path or slowest motion that would explain the image data.

We formalize this preference for slow speeds using a prior probability on the two dimensional space of velocities that is Gaussian and centered on the origin. According to this prior, in the absence of any image data, the most probable velocity is zero (i.e., no motion), and slower velocities are generally more likely to occur than fast ones. As with the noise model, we have no direct evidence (either from first principles or from empirical measurements) that this assumption is correct. But we show in the next section that it is sufficient to account qualitatively for much of the perceptual data.

Under the Bayesian framework, the percept of the ideal observer is based on the posterior probability (the probability of a velocity given the image measurements), which is computed from the likelihood and prior using Bayes' rule (see **Methods**). In our case, the posterior is formed by multiplying the prior and the likelihoods at all image locations. This is correct under the assumptions that the noise in the measurements is statistically independent, and that the likelihoods being multiplied correspond to image locations that are moving at the same velocity.

One can calculate the velocity estimate of the ideal observer as the mean or the maximum of the posterior distribution. In our case, the posterior distribution is Gaussian, and the mean (which is equal to the maximum) velocity may be computed analytically using the following matrix equation:

$$
v^* = - \begin{pmatrix} \sum I_x^2 + \frac{\sigma^2}{\sigma_p^2} & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 + \frac{\sigma^2}{\sigma_p^2} \end{pmatrix}^{-1} \begin{pmatrix} \sum I_x I_t \\ \sum I_y I_t \end{pmatrix} \tag{1}
$$

where I_x, I_y, I_t refer to the spatial and temporal derivatives of the image sequence. The sums are taken over all locations that translate together (in this paper, we assume this includes the entire image). Note that this equation allows us to obtain a prediction of the ideal observer for any image sequence. Note also that the solution of Eq. (1) has only one free parameter: the ratio of σ to σ_p. Changing both of these while holding the ratio fixed changes the width of the posterior but not its peak.

Figs. 4a-c illustrate the calculation of the posterior for the moving rhombus stimuli.

The free parameter (σ/σ_p) is held constant for the simulations. Consistent with the human percept, the ideal observer predicts horizontal motion for a high contrast narrow rhombus, diagonal motion for a low contrast narrow rhombus and nearly horizontal motion for a low contrast fat rhombus. Fig 4d shows a more quantitative comparison of the ideal observer and the perceived motion of rhombuses. We showed three subjects a continuum of low-contrast rhombuses that varied between the extremes of 'thin' and 'fat' and asked them to indicate the perceived direction by positioning the cursor of a computer mouse. Each experiment consisted of 100 presentations. The circles in Fig. 4d show the results for a single naive subject (all subjects showed a similar effect). The solid line shows the predictions of Eq. (1), which provides an excellent fit to the experimental data.

In order to illustrate the effect of the free parameter, the dotted lines in Fig. 4d indicate the predictions when the free parameter is decreased by a factor of 10 (top dotted line) or increased by a factor of 10 (bottom line). Note that over two orders of magnitude of change in this parameter, the qualitative predictions remain unchanged. In fact, there is no setting of the free parameter that would make the perception of narrow rhombuses more veridical than fat ones. Similarly, there is no setting that would make the perception of low contrast rhombuses more veridical than high contrast ones.

Results

We compared the predictions of the ideal observer (i.e., the solution of Eq. (1)) to previously published psychophysical data. The free parameter was hand-adjusted for each experiment, but held constant for all conditions within each experiment. One might expect that different observers make different "assumptions" regarding noise, and indeed significant individual differences between observers have been reported in the magnitude of these illusions[17]. As with the rhombus example, the value of the free parameter does not change the qualitative predictions of the model for any of the stimuli discussed here.
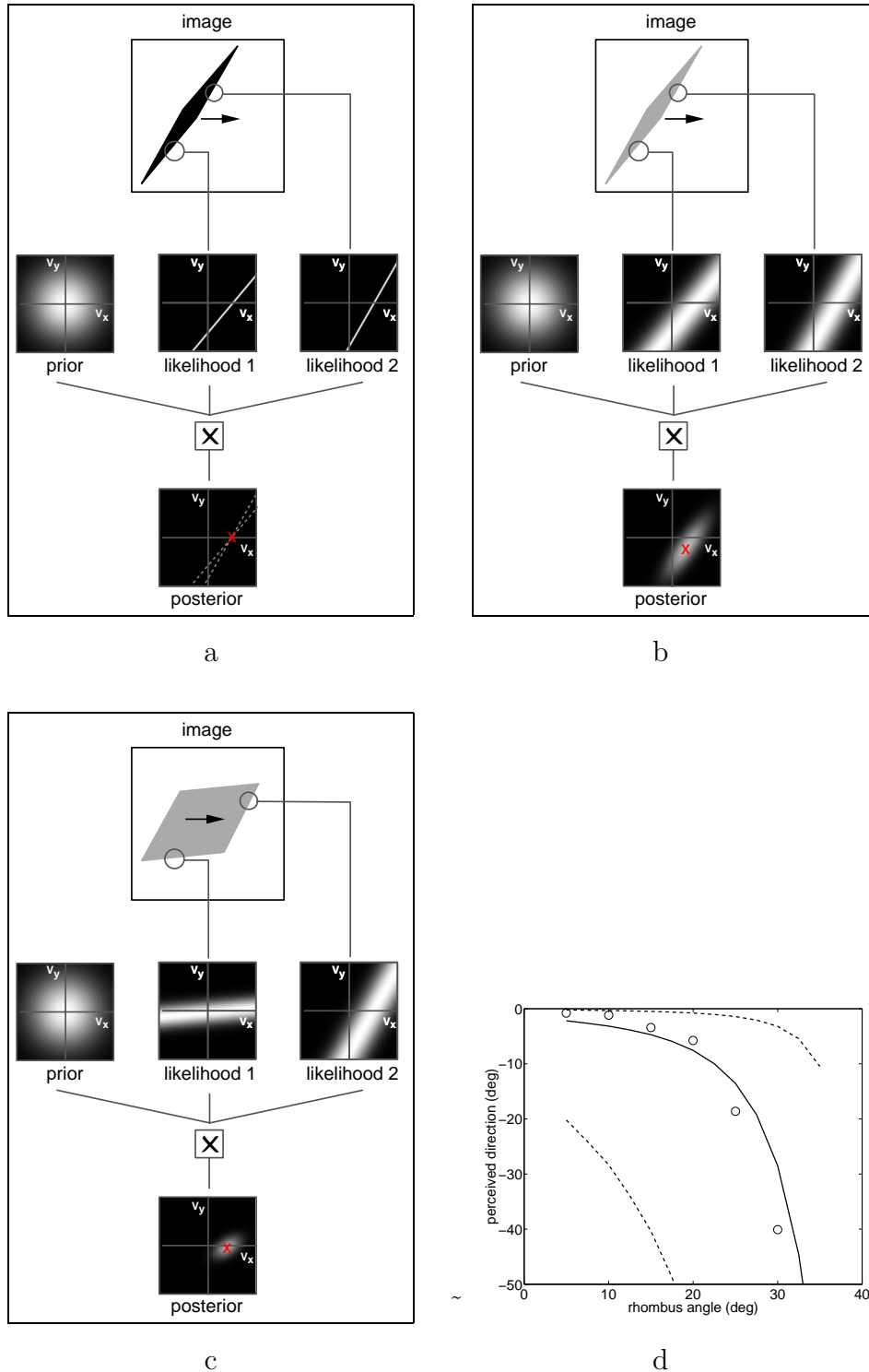
Fig. 4: Predictions of ideal observer for rhombus stimuli. **a-c** Construction of the posterior distribution for the rhombus stimuli. For illustration purposes, we show likelihood functions for only two locations; the estimator used elsewhere in the paper incorporates likelihoods from all locations. **d.** Circles show perceived direction for a single observer as rhombus angle is varied (a gradual shift from thin to fat rhombuses). Solid line shows the predictions of the Bayesian estimator, computed using Eq. (1) where the free parameter was varied by hand to fit the data. Dotted lines show the effect of varying the free parameter over two orders of magnitude.
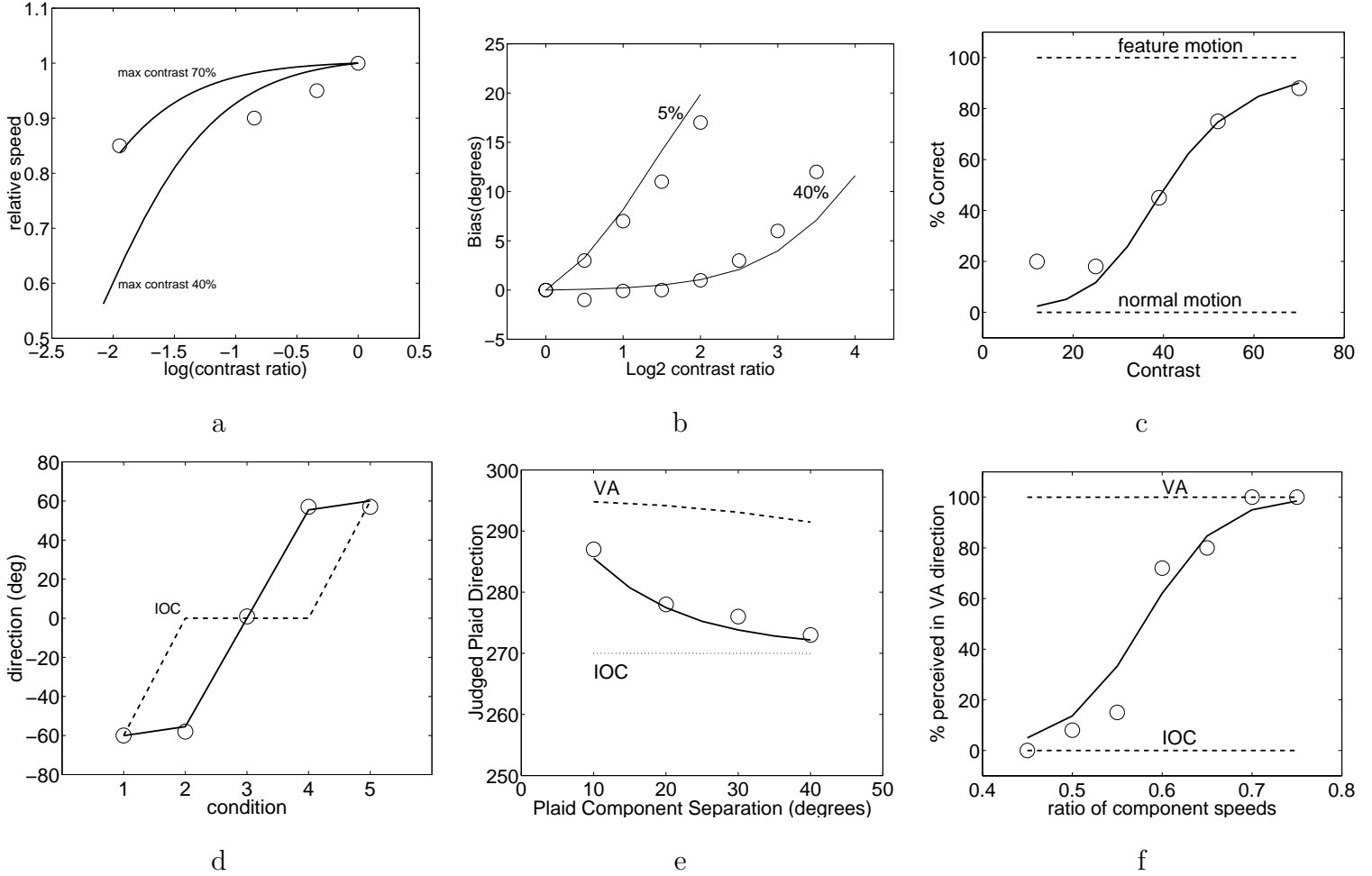
Fig. 5: Comparison of ideal observer (solid line) to a variety of published psychophysical data (circles). **a.** Contrast influence on perceived grating speed[31]. **b.** Relative contrast influence on perceived plaid direction[20]. **c.** Contrast influence on perceived line direction[32]. **d.** Perceived direction of Type I versus Type II plaids[17]. **e.** Relative orientation influence on Type II plaids[18]. **f.** Relative speed influence on type II plaids[19].

Contrast influence on perceived grating speed. The perceived speed of a single grating depends on contrast[33, 31, 34, 35], with lower-contrast patterns consistently appearing slower [34]. This tendency has been implicated in the tendency of automobile drivers to speed up in the fog [36]. Figure 5a shows the results of a psychophysical experiment quantifying this effect[31]. Subjects were asked to compare the apparent speed of two gratings of different contrast. The low contrast grating is consistently perceived as moving slower, and the effect depended primarily on the ratio of contrasts of the two gratings (i.e., it was approximately independent of absolute contrast). The circles in the graph of Fig. 5a indicate the speed of the lower-contrast grating relative to the higher-contrast grating, as a function of the contrast ratio. The perceived speed is an approximately linear function of the contrast ratio.

The ideal observer shows a similar contrast dependence. At low contrasts, the likelihood is broader and the prior has a stronger influence on the estimate. The two lines in Fig. 5a also show the relative speeds as a function of contrast ratio for two different maximal contrasts. Consistent with the human percept, the ideal observer estimates the low contrast grating as moving slower.

The simple ideal observer presented here does not predict the quasilinear shape of the perceived relative speeds, nor does it predict the lack of dependence on total contrast: the predictions when the higher contrast is 40% and 70% are slightly different. In the discussion section, we describe the construction of a slightly more elaborate model that can account more quantitatively for these effects.

Relative contrast influence on perceived plaid direction. The perceived direction of a plaid depends on the relative contrast of the two constituent gratings[20]. Fig. 5b shows the results of an experiment in which subjects reported the perceived direction of motion of symmetric plaids while the contrast ratio of the two components was varied. Perceived direction is always biased toward the normal direction of the higher contrast grating. The magnitude of the bias changes as a function of the total contrast of the plaid (i.e., the sum of the contrasts of the two gratings). When the contrast of both gratings is increased

12

(while holding the ratio of contrasts fixed) a smaller bias is observed. The ideal observer shows a similar effect(E. P. Simoncelli & D. J. Heeger, *Invest. Opthal. Vis. Sci. Suppl.* **33**, 954,1992), which again follows from the fact that at low contrast, there is higher uncertainty and hence the low-contrast grating has less influence on the estimate.

Contrast influence on perceived line direction. Subjects tend to misperceive the direction of a moving line at low contrasts, even when its endpoints are visible [32]. Fig. 5c shows the results of an experiment in which subjects were asked to report the direction of a 'matrix' of lines. The matrix was constructed by replicating a single line at multiple locations in the visual field. The line was oriented such that even when the line was moving upward the normal velocity was downward. Subjects were asked to judge whether the matrix was moving up or down. Note that at low contrasts, performance is far below chance indicating subjects perceived upward motion while the line actually moved downward. The authors modeled these results using two separate mechanisms, one dealing with terminator motion and other with line motion. The terminator mechanism was assumed to be active primarily at high contrast and the line mechanism at low contrast.

The ideal observer also misperceives the direction of motion for this stimulus at low contrast. At low contrast, the likelihoods are broader and the estimator prefers the normal velocity (which is slower than the true velocity). To obtain a percent correct for the ideal observer we assumed that v^* is corrupted by decision noise and calculated the probability that the corrupted v^* was upward. The decision noise was Gaussian in velocity space. The standard deviation of the decision noise determines the sharpness of the psychometric function and was adjusted by hand. The solid line in Fig. 5d shows the predicted percent correct for the ideal observer: it agrees with the human percept.

Perceived direction of Type I versus Type II plaids. In the plaid literature, a distinction is often made between two types of configuration: for a 'Type I' plaid, the direction of the veridical velocity lies between that of the two normal velocities, while for a 'Type II' plaid the veridical direction lies outside the two normals[17]. In the latter case,

the vector average is quite different from the veridical velocity.

At low contrast, the perceived direction for Type II plaids is strongly biased in the direction of the vector average while the perceived direction of Type I plaids is largely veridical. Figure 5d plots the perceived direction of a plaid under five different conditions for a single subject[17]. In all five conditions, the angular separation between the two gratings is 22.3 degrees, but in conditions 1, 3, 5 the two normal velocities are on different sides of the veridical motion (Type I) while in conditions 2, 4 they are on the same side of the veridical motion (Type II). The dotted line shows the prediction of IOC: as can be seen, the Type I plaids are perceived to move in the IOC direction but the Type II plaids are seen to move at approximately the VA direction, which is approximately 55 degrees away from the veridical motion. The authors explained their findings using a contrast-dependent combination of first order and second order motion analyzers[37].

The ideal observer also predicts different directions of motion for the two types of plaids at low contrast. The solid line shows the predicted direction. The 'misperception' of Type II plaids is similar to the situation with the narrow rhombus: the VA velocity is significantly slower than the IOC solution and hence it is favored at low contrasts. This bias toward the VA decreases in the ideal observer with increasing contrast, as the likelihoods become narrower.

It has also been reported that the VA bias is more pronounced with shorter presentation durations[17]. Our ideal observer is based on instantaneous measurements and does not depend on the display duration. It is straightforward to extend the formulation so that the ideal observer integrates information over time. In this case, increased duration acts in a similar fashion to increased contrast: the longer the duration the narrower the likelihood. Such an extended formulation would predict that the VA bias would decrease with increased duration. Lorenceau et al. found a similar effect of duration in their stimuli, and this would also be predicted by the same extension of our model.

Relative orientation influence on Type II plaids. The perceived direction of a Type II plaid depends strongly on the angle between the components[18]. Figure 5e plots

the perceived direction as a function of the angle between the components (the plaid pattern velocity was held constant). The perceived direction is inconsistent with a pure VA mechanism or a pure IOC mechanism. Rather it shows a gradual shift from the VA to the IOC solution as the angle between the components increases. The solid line shows the prediction of the ideal observer (the direction of v^* in Eq. 1). The situation is similar to the 'narrow' versus 'fat' rhombuses (Fig. 4). When two likelihoods whose constraint lines are nearly identical are multiplied, their product will be broad and hence have less of an influence on the posterior. Compare this to a situation where the two likelihoods have widely differing constraint lines: now the product of the two will be narrow and hence have greater influence on the posterior.

Relative speed influence on type II plaids. The perceived direction of a plaid also depends on the relative speeds of the components[19]. Figure 5f shows data for a single subject who viewed a plaid with IOC and VA directions on either side of vertical, and was asked whether or not the motion was left or right of vertical. When the speeds of the two components were similar, the subject answered rightwards (consistent with the VA solution), but when the speeds were dissimilar, the subject answered leftwards (consistent with the IOC solution). We found that the ideal observer (Eq. 1) shows a similar shift from leftward to rightward velocities. We again calculated a 'percent correct' value for the ideal observer by assuming decision noise. The solid line in Fig. 5f shows the predictions.

Discussion

Research on visual motion analysis has yielded a tremendous amount of experimental data. When viewed in the context of existing rules such as IOC and VA, these data seem contradictory and appear to require an arbitrary combination scheme that applies the right rule in the right conditions. Such an approach can successfully fit the data, but is typically lacking in predictive power: with a complicated enough combination scheme

15

one can model any experiment. More importantly, since these rules are not formulated directly on image measurements, it is not clear how one should generalize them for application to arbitrary spatiotemporal stimuli.

In this paper we have taken an alternative approach. We derived an optimal estimator for local image velocity using the standard assumption of intensity constancy and two additional assumptions: measurement noise and and an a priori preference for slower velocities. We find that, like humans, the motion estimates of this model include apparent biases and illusions. Moreover, the predicted non-veridical percept is quite similar to that exhibited by humans under the same circumstances. While the model does not account for all of the existing data quantitatively, we were surprised by the wide range of effects predicted by these simple assumptions.

As mentioned in **Results**, our model does not provide a good quantitative fit to the data of Fig. 5a, which suggest a quasilinear dependence of perceived grating speed on contrast, and minimal dependence on total contrast. Hurlimann et al. (F. Hurlimann & D. Kiper & M. Carandini *Invest. Opthal. Vis. Sci. Suppl.* **40**, 794, 2000) extended our model by including a nonlinear 'gain control' function to map stimulus contrast into perceived contrast. For each subject in their study, they measured a gain control function from contrast discrimination experiments. They then used the perceived contrast rather than stimulus contrast as input to our model. They found that "when a realistic representation of contrast is introduced" the quantitative predictions of the Bayesian model "are in general agreement with the data".

In further simulations that are available at http://neurosci.nature.com/web_specials, we asked: is there a monotonic nonlinear gain control function that will enable our model to better fit the results in [31] ? We used a numerical search procedure to find such a function obtained an excellent fit to the results in [31] as well as continuing to fit all other data mentioned in this paper.

One result that is not predicted by our model is the finding that low contrast gratings actually appear to move faster than high contrast gratings for temporal frequencies above 8Hz [33] . A later paper by the same author, however, failed to reproduce this reversal

result using a forced choice paradigm [31], and concluded that "the apparent reversal found previously is therefore probably an artifact of the experimental method with subjects making 'speed' matches based on some other criterion".

Our Bayesian estimator is meant as a perceptual model, and does not specify a particular implementation. Nevertheless, the solution can be instantiated using so-called motion energy mechanisms [28, 38], and detailed models of the physiology of the motion pathway [28, 39, 40, 41, 24, 25] suggest that a population of MT cells may be forming a representation of the local likelihood of velocity. In addition, we believe it should be possible to refine and justify the assumptions we've made. In particular, the prior distribution on velocity could be estimated empirically from the statistics of motion in the world. In a physiological implementation, the noise model should be replaced by one that more accurately reflects the uncertainties of neural responses.

Our model also suggests some novel experiments. First, if the single free parameter is observer-dependent (but otherwise constant), the magnitude of different illusions for the same subject should be correlated (e.g. observers who greatly underestimate the speed of low contrast gratings should also show a larger bias towards VA in type II plaids). Second, in all our simulations we have used only the maximum (or mean) of the posterior distribution. It would be interesting to devise psychophysical experiments that test whether human percepts reflect the shape of the full posterior distribution.

We have focused on an ideal observer for estimating a single two-dimensional translation. This model cannot estimate more complicated motions such as rotations and expansions, nor can it handle scenes containing multiple motions. In [29] we describe an extended ideal observer for more general scenes with multiple motions. We show that an ideal observer that assumes that velocity fields should be 'slow and smooth' [42] can explain an even wider range of motion phenomena. In particular, in (H. Farid & E. P. Simoncelli, *Invest. Opthal. Vis. Sci. Suppl.* **35**, 1271, 1994)[29] we discuss how the bias toward slower motions can sometimes account for one of the most critical issues in motion perception: the question of whether to combine measurements into a single coherent motion or assume that there are actually multiple motions.

While the details of our model should certainly be refined and extended to handle more complicated phenomena, we believe the underlying principle will continue to hold: that many motion 'illusions' are not the result of sloppy computation by various components in the visual system, but rather a result of a coherent computational strategy that is optimal under reasonable assumptions.

Methods

Most models of early motion extraction rely on an assumption of 'intensity conservation'. Under this assumption, the points in the world, as measured in the image, move but do not change their intensity over time. Mathematically, this is expressed as:

$$I(x, y, t) = I(x + v_x \Delta t, y + v_y \Delta t, t + \Delta t), \tag{2}$$

where v_x and v_y are the components of the vector, v, describing the image velocity. If we assume that the observed image is noisy, then intensity is not conserved exactly. Thus, Eq. (2) becomes

$$I(x, y, t) = I(x + v_x \Delta t, y + v_y \Delta t, t + \Delta t) + \eta, \tag{3}$$

where η is a random variable representing noise.

We want to use Eq. (3) to derive the likelihood at location i, $P(I(x_i, y_i, t)|v_i)$. This requires additional assumptions. We assume the noise, η, is Gaussian with standard deviation σ. We further assume that the velocity is constant in a small window around x_i, y_i and that the intensity surface $I(x, y, t)$ is sufficiently smooth so it can be approximated by a linear function for small temporal durations. We can thus replace $I(x + v_x \Delta t, y + v_y \Delta t, t + \Delta t)$ with its first-order Taylor series expansion, which gives:

$$P(I(x_i, y_i, t)|v_i) \propto \tag{4}$$
$$\exp\left(-\frac{1}{2\sigma^2} \int_{x,y} w_i(x, y)(I_x(x, y, t)v_x + I_y(x, y, t)v_y + I_t(x, y, t))^2 \, dx \, dy\right),$$

where $\{I_x, I_y, I_t\}$ denote the spatial and temporal derivatives of the intensity function I, and $w_i(x, y)$ is a window centered on (x_i, y_i). The likelihoods shown in Figs. 3 and 4 are computed from Eq. (4) with $w(x, y)$ a small Gaussian window.

Finally, we assume a prior favoring slow speeds:

$$P(v) \propto \exp\left(-\|v\|^2/2\sigma_p^2\right). \tag{5}$$

The posterior probability of a velocity is computed by combining the likelihood and prior using Bayes' rule. Because we assume the noise is independent over spatial location, the total likelihood function is just a product of likelihoods:

$$P(v|I) \propto P(v) \prod_i P(I(x_i, y_i)|v), \tag{6}$$

where the product is taken over all locations i that are moving with a common velocity (i.e., $v_i = v$). Substituting Eqs. (4) and (5) into Eq. (6) gives:

$$P(v|I) \propto \exp\left(-\|v\|^2/2\sigma_p^2 - \frac{1}{2\sigma^2}\int_{x,y}\sum_i w_i(x,y)(I_x(x,y)v_x + I_y(x,y)v_y + I_t)^2 dxdy\right).$$

In this paper, we assume the entire image moves according to a single translational velocity, and so sum over all spatial positions. In this case, $\sum_i w_i(x,y)$ is a constant so the posterior probability is given by:

$$P(v|I) \propto \exp\left(-\|v\|^2/2\sigma_p^2 - \frac{1}{2\sigma^2}\int_{x,y}(I_x(x,y)v_x + I_y(x,y)v_y + I_t)^2 dxdy\right).$$

To find the most probable velocity, we replace the integral with a discrete sum, take the logarithm of the posterior, differentiate it with respect to v and set the derivative equal to zero. The logarithm of the posterior is quadratic in v so that the solution can be written in closed form using standard linear algebra. The result is given in Eq. (1).

Acknowledgements

References

1. Nakayama, K. Biological image motion processing: A review. *Vision Research* **25**, 625–660 (1985).

2. Horn, B. K. P. and Schunck, B. G. Determining optical flow. *Artif. Intell.* **17**(1–3), 185–203, August (1981).

3. Lucas, B. D. and Kanade, T. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th Int'l Joint Conf on Artificial Intelligence*, 674–679 (, Vancouver, 1981).

4. Wuerger, S., Shapley, R., and Rubin, N. On the visually perceived direction of motion by Hans Wallach: 60 years later. *Perception* **25**, 1317–1367 (1996).

5. Wallach, H. Ueber visuell whargenommene bewegungrichtung. *Psychologische Forschung* **20**, 325–380 (1935).

6. Marr, D. and Ullman, S. Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London B* **211**, 151–180 (1981).

7. Adelson, E. and Movshon, J. Phenomenal coherence of moving visual patterns. *Nature* **300**, 523–525 (1982).

8. Movshon, A., Adelson, E., Gizzi, M., and Newsome, W. The analysis of moving visual patterns. *Experimental Brain Research* **11**, 117–152 (1986).

9. Welch, L. The perception of moving plaids reveals two processing stages. *Nature* **337**, 734–736 (1989).

10. Morgan, M. Spatial filtering precedes motion detection. *Nature* **355**(6358), 344–6 (1992).

11. Schrater, P., Knill, D., and Simoncelli, E. Mechanisms of visual motion detection. *Nature Neuroscience* **3**(1), 64–8 (2000).

12. Rodman, H. and Albright, T. Single-unit analysis of pattern motion selective properties in the middle temporal visual area MT. *Experimental Brain Research* **75**, 53–64 (1989).

13. Movshon, J. A. and Newsome, W. T. Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. *Visual Neuroscience* **16**(23), 7733–7741 (1996).

14. Okamoto, H., Kawakami, S., Saito, H., Hida, E., Odajima, K., Tamanoi, D., and Ohno, H. MT neurons in the macaque exhibited two types of bimodal direction tuning as predicted by a model for visual motion detection. *Vision Research* **39**(20), 3465–3479 (1999).

15. Ferrera, V. and Wilson, H. Perceived direction of moving two-dimensional patterns. *Vision Research* **30**, 273–287 (1990).

16. Mingolla, E., Todd, J., and Norman, J. The perception of globally coherent motion. *Vision Research* **32**(6), 1015–1031 (1992).

17. Yo, C. and Wilson, H. Perceived direction of moving two-dimensional patterns depends on duration, contrast, and eccentricity. *Vision Research* **32**(1), 135–147 (1992).

18. Burke, D. and Wenderoth, P. The effect of interactions between one-dimensional component gratings on two dimensional motion perception. *Vision Research* **33**(3), 343–350 (1993).

19. Bowns, L. Evidence for a feature tracking explanation of why type II plaids move in the vector sum directions at short durations. *Vision Research* **36**(22), 3685–3694 (1996).

20. Stone, L., Watson, A., and Mulligan, J. Effect of contrast on the perceived direction of a moving plaid. *Vision Research* **30**(7), 1049–1067 (1990).

21. Rubin, N. and Hochstein, S. Isolating the effect of one-dimensional motion signals on the perceived direction of moving two-dimensional objects. *Vision Research* **33**, 1385–1396 (1993).

22. Helmholtz, H. *Treatise on Physiological Optics.* Thoemmes, (2000). orig. publication 1866.

23. Knill, D. and Richards, W. *Perception as Bayesian Inference.* Cambridge University Press, (1996).

24. Ascher, D. and Grzywacz, N. A Bayesian model for the measurement of visual velocity. *Vision Research* **40**, 3427–3434 (2000).

25. Koechlin, E., Anton, J. L., and Burnod, Y. Bayesian inference in populations of cortical neurons: a model of motion integration and segmentation in area mt. *Biol. Cybernetics* **80**, 25–44 (1999).

26. Simoncelli, E., Adelson, E., and Heeger, D. Probability distributions of optical flow. In *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, 310–315, (1991).

27. Heeger, D. J. and Simoncelli, E. P. Model of visual motion sensing. In Spatial Vision in Humans and Robots, Harris, L. and Jenkin, M., editors, chapter 19, 367–392. Cambridge University Press (1994).

28. Simoncelli, E. P. *Distributed Representation and Analysis of Visual Motion.* PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts of Technology, Cambridge, , January (1993).

29. Weiss, Y. *Bayesian motion estimation and segmentation.* PhD thesis, Department of Brain and Cognitive Sciences, Massachusetts of Technology, Cambridge, , May (1998).

30. Ullman, S. *The interpretation of visual motion.* The MIT Press, (1979).

31. Stone, L. and Thompson, P. Human speed perception is contrast dependent. *Vision Research* **32**, 1535–1549 (1990).

32. Lorenceau, J., Shiffrar, M., Wells, N., and Castet, E. Different motion sensitive units are involved in recovering the direction of moving lines. *Vision Research* **33**(9), 1207–1217 (1992).

33. Thompson, P. Perceived rate of movement depends on contrast. *Vision Research* **22**, 377–380 (1982).

34. Thompson, P., Stone, L., and Swash, S. Speed estimates from grating patches are not contrast normalized. *Vision Research* **36**(5), 667–674 (1996).

35. Blakemore, M. and Snowden, R. The effect of contrast upon perceived speed: a general phenomenon? *Perception* **28**(1), 33–48 (1999).

36. Snowden, R., N. Stimpson, N., and Ruddle, S. Speed perception fogs up as visibility drops. *Nature* **392**(6675), 450 (1998).

37. Wilson, H., Ferrera, V., and Yo, C. A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience* **9**, 79–97 (1992).

38. Weiss, Y. and Fleet, D. Velocity likelihoods in biological and machine vision. In Probabilistic models of the brain, R.Rao, Olshausen, B., and Lewicky, M., editors. MIT Press, (2002).

39. Nowlan, S. J. and Sejnowski, T. J. A selection model for motion processing in area MT of primates. *The Journal of Neuroscience* **15**(2), 1195–1214 (1995).

40. Simoncelli, E. and Heeger, D. A model of neuronal responses in visual area MT. *Vision Research* **38**(5), 743–761 (1998).

41. Pouget, A., Dayan, P., and Zemel, R. Information processing with population codes. *Nat Rev Neurscience* **1**(2), 125–32 (2000).

42. Grzywacz, N. and Yuille, A. Theories for the visual perception of local velocity and coherent motion. In Computational models of visual processing, Landy, J. and Movshon, J., editors. MIT Press, Cambridge, Massachusetts (1991).