Illusory motion due to causal time filtering

Cornelia Fermüller [a 1], Ji Hui [b] and Akiyoshi Kitaoka [c]

[a] *Computer Vision Laboratory*
Center for Automation Research
Institute for Advanced Computer Studies
University of Maryland
College Park, MD 20742-3275
email: fer@cfar.umd.edu

[b] *Department of Mathematics*
National University of Singapore
Singapore

[c] *Department of Psychology*
Ritsumeikan University
Japan

Abstract

A new class of patterns, composed of repeating patches of asymmetric intensity profile, elicit strong perception of illusory motion. We propose as main cause for this illusion erroneous estimation of image motion induced by fixational eye movements. Image motion is estimated with spatial and temporal energy filters, which are symmetric in space, but asymmetric (causal) in time. That is, only the past, but not the future, is used to estimate the temporal energy. It is shown that such filters mis-estimate the motion of locally asymmetric intensity signals at certain spatial frequencies. In an experiment the perception of the different illusory signals was quantitatively compared by nulling the illusory motion with opposing real motion, and was found to be predicted well by the model.

Key words: illusory motion, spatio-temporal filtering, image motion estimate, computational model

[1] Corresponding author

Fig. 1. Variation of Donguri pattern. In peripheral vision most observer experience rotary movement. The direction in the circular arrangements alternates, with counter-clockwise direction in the upper left.

1 Introduction

Most observers experience very strong illusory movement when viewing patterns such as Donguri (Fig.1) and Rotating Snake (Fig.2) (Kitaoka, 2003). These patterns are composed of image patches which have an asymmetric intensity profile. For example, consider a narrow slice in the middle region of one of the ovals in Donguri as shown in Fig. 3a. (The Japanese word "donguri" translates to "acorn".) Its mono-chromatic intensity image can be described as a white and a dark bar (the boundaries of the oval) next to different shades of gray. Referring to Fig.3b, from the highest intensity (the white bar) the intensity drops about twice as much on the right than on the left side. Similarly, from the lowest intensity (the dark bar) the intensity rises about twice as much as on the right than on the left. Thus, at the two bars the change of intensity in the right and left neighborhood is different. Informally we say that the pattern is asymmetric. Patterns with such intensity profiles create a very strong illusory effect. The perceived movement is a drift from the intensity extremum in the direction of lesser intensity change (i.e. from the white bar to light gray, and from the dark bar to dark gray) (Kitaoka and Ashida, 2003).

The illusory movement requires that one moves the eyes, and it is perceived in non-central vision. It stops if steadily fixating after about 6 - 8 seconds.

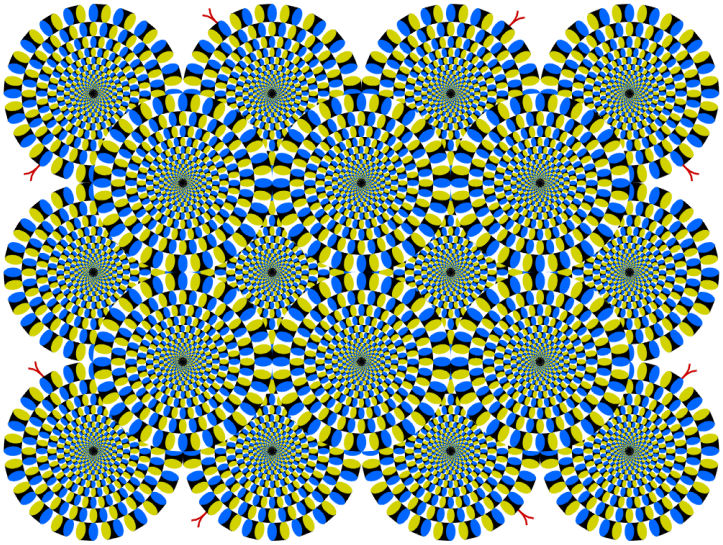


Fig. 2. Rotating Snake.

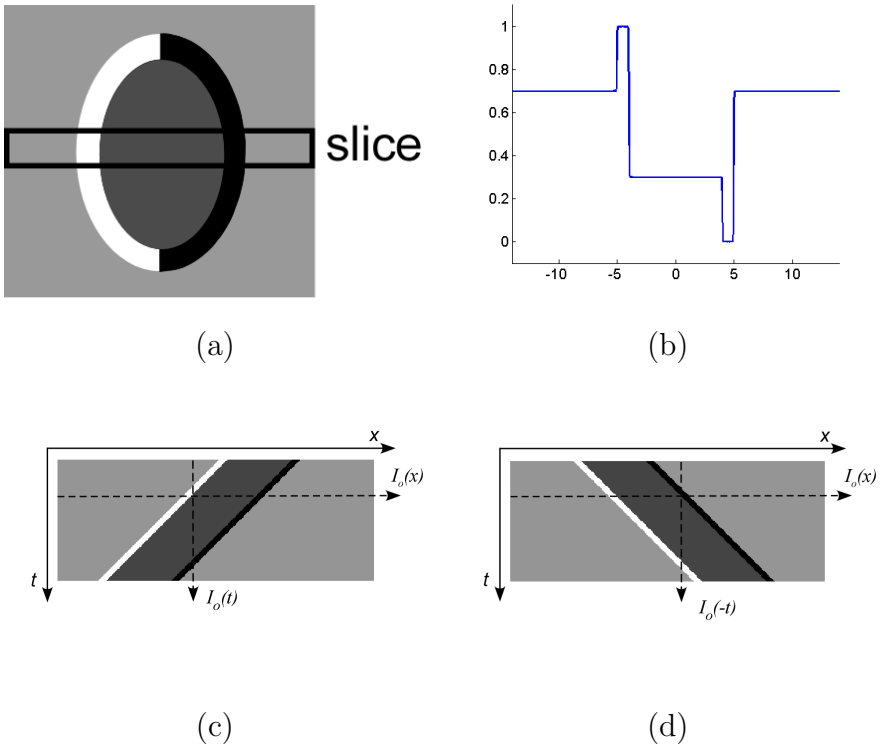


Fig. 3. (a) Slice through a patch in the Donguri pattern. (b) Its intensity profile. (c,d) Spatio-temporal picture of the patch moving to the left and to the right. $I_0(x)$ denotes a static image and $I_0(t)$ denotes the signal at a point over time.

The perceived motion is a drift, whose direction depends on the intensity relationship of the pattern elements. Chromaticity is not necessary for the illusion, but enhances the effect in some patterns (Kitaoka, 2006). The illusion depends on the size of the image patterns. For medium sized patches such as

Donguri, motion occurs in a patch when it is viewed in the periphery. Smaller patches give illusory motion closer to the center of the retina. Blur reduces the illusion in peripheral vision, but enhances it in central vision. The illusory effect is more forceful if a pattern consists of many patches, and the patches are at multiple sizes. It is stronger when the patches are circularly organized, but also exists for columnar and other arrangements.

It is generally considered that the illusory effect was first observed in patterns with circularly organized sawtooth luminance grating named the staircase illusion (Fraser and K.J.Wilcox, 1979) and the peripheral drift illusion (J. Faubert and Herbert, 1999) (see Fig 18). Ashida and Kitaoka (2003) showed that the effect is much increased if the sawtooth luminance profile is replaced by step functions with intensities in the same order as in Donguri (i.e. light gray - white - dark gray - black), and if the large patches are replaced by many small ones. In Kitaoka and Ashida (2004) the authors presented patterns with continuously increasing intensity ramp-like profiles, which are perceived in central and close to central vision, and Kitaoka in (2006) proposed a classification of the different intensity profiles giving rise to the illusory effect.

A number of hypotheses for the illusory motion have been proposed. The dominant idea originating from J. Faubert and Herbert (1999) is that temporal differences in luminance processing produce a signal that tricks the motion system. The theories differ in how this signal is produced. Faubert and Herbert suggest that eye movements or blinks need to trigger an image motion, and the different motion signals (due to differences in intensity) are integrated over large spatial areas. Bacus and Oruç (1995) focus on the perception during fixation and hypothesize that motion is not necessary, but a motion signal is triggered from the change of the neural response over time. Differently strong contrasts and intensities cause different neural response curves over time (Albrecht et al., 2002). As a result the phase of the signal is estimated erroneously as time passes and a motion signal is triggered. Using psychophysical and physiological experiments with stimuli consisting of pattern frames interlaced with frames of the pattern's average intensity Conway et al. (2005) argue that in addition to signals created by the differences in intensity processing, a signal analogous to the reverse phi motion (Anstis, 1970) is created. Phi motion is an image motion effect caused by inverting the intensities in some frames of a video sequence. It is easy to explain that such motion will arise under the conditions described. However, the illusory effect under free viewing and the effect when flickering the pattern do not appear to be the same. The latter is experienced much stronger for the reduced experimental stimuli and is experienced even by observers who do not have the effect under free viewing.

We propose that the main reason for the illusion is erroneous estimation of the image motion due to involuntary fixational eye movements. Work by Mu-

rakami et al. (2006) indicates that these eye movements are the drifts.[2]. In particular, the authors showed a correlation between the amplitude of drift movements in different observers and their illusory perception. The small eye movements cause a change of the image on the retina and trigger the estimation of a motion field. This motion field is due to rigid motion and thus has a certain structure. Under normal circumstances the vision system estimates this image motion and compensates for it, i.e. the images are stabilized (Murakami, 2004; Murakami and Cavanagh, 1998). Even for asymmetric signals, the vision system estimates the correct 3D rigid motion using the average of all the motion vectors in the patterns. However, a mis-estimation occurs at certain locations in the image. The difference between the estimated rigid motion field and the erroneously estimated image motion vectors gives rise to residual motion vectors. These residual motion vectors, which are integrated over time, then cause the perception of illusory motion in the image.

The dominant model for motion processing in humans and other mammals is the motion energy model (Watson and Ahumada, 1985; Adelson and Bergen, 1985), and it has been found to be consistent with the physiological responses in primary visual cortex (Albrecht and Geisler, 1991). Motion is found from the response of multiple spatio-temporal filters, which are separable in space and time. The spatial filters are symmetric. The temporal filters, however, are asymmetric. In other words, temporal filtering is causal. The filters receive as input data from the present and the past; but the processing does not wait to collect data from the future. A restriction to causal filters is natural for real time systems; for a system to be non-causal, it would have to collect data from the future in order to make an estimate about the present state. Due to the causal filtering, the image motion will be mis-estimated in asymmetric image signals at certain spatial frequencies. The resolution of the eye decreases from the center to the periphery, and with it the frequency of the energy filters decreases. An image patch of certain size will thus be estimated of erroneous motion only on some parts of the eye. For most image patches the mis-estimation occurs at relatively low frequencies, and mis-perception occurs on the periphery. For very small patterns the perception is closer to the center.

The next two sections will explain in detail the reasons for mis-estimation of the image motion. The main concept, however, can be easily explained. Figure 4 illustrates a spatio-temporal filter with symmetric impulse response in the spatial domain, and with asymmetric impulse response (Burr and Morrone, 1993) in the time domain. The spatial response may be modeled as a sinusoid of certain frequency enveloped by a symmetric function. The temporal response

[2] The drift movements, one of the three fixational eye movements, are defined as incessant random fluctuations at about 1-30 Hz, quite large ($\sim$ 10 min of visual angle) and fast (up to $2-3^{\circ}/sec$) (Eizenman et al., 1985). They are more rigorous after a saccade (Ross et al., 2001) than during steady fixation.

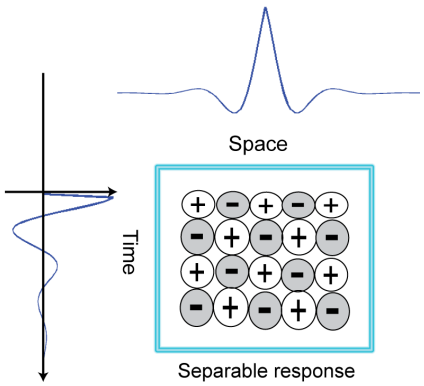


Fig. 4. Illustration of biological implementation of spatio-temporal filter (similar to Fig.6 of (Adelson and Bergen, 1985)). The spatial and temporal impulse responses are shown along the margins. Their product is shown schematically in the center.

may be modeled as the same sinusoid enveloped by an asymmetric function. Consider filtering the Donguri signal with a whole range of spatial filters of increasing size (and decreasing frequency). Let's go ahead in the paper and take a look at Figures 7b-d, which show the amplitude of the response from filtering a single bar in Donguri. A filter of high spatial frequency will respond to the two edges bordering the bar. A filter of low spatial frequency will not recognize the edges, but only have one response to the bar. However, for intermediate frequencies, with the period of the sinusoid about as large as the bar, the two edges will effect each other during filtering leading to a response curve of two merged peaks. As a result the filter will poorly localize the spatial frequencies. Because the temporal filters are asymmetric as well, for these spatial frequencies the temporal frequency responses and thus the estimated image motion from a movement to the left and a movement to the right will be significantly different. There is larger estimated left motion than right motion.

The idea of anisotropic temporal filtering was first proposed by Ashida and Kitaoka (2003), who modeled image motion estimation using a differential motion model with asymmetric temporal derivative filters. This local model, however, requires the filters to have larger weight in the past and smaller weight in the present. Furthermore, it cannot explain the estimation at different resolutions of the pattern. For this we need to look at the different frequencies.

2 Motion Estimation in the Frequency Domain

The monochromatic light distribution on the retina can be described as a function $I(x, y, t)$, which specifies the intensity at a point (x, y) at time t. We

refer to the instantaneous light distribution at time $t = 0$ as the static image $I_0(x, y) = I(x, y, 0)$. Let us assume that within a small interval the change of the image can be described as a translation with constant motion velocity $\vec{v}$ of horizontal and vertical speed components (u, v). Thus, the intensity functions at time t and at time 0 are related as

$$I(x, y, t) = I(x - ut, y - vt, 0). \tag{1}$$

Taking the three-dimensional Fourier transform of this equation, we obtain

$$\mathcal{F}\{I(x, y, t)\}(\omega_x, \omega_y, \omega_t) = \mathcal{F}\{I(x, y, 0)\}(\omega_x, \omega_y, \omega_t) \cdot \delta(u\omega_x + v\omega_y + \omega_t), \tag{2}$$

where ω_x, ω_y denote the spatial frequencies and ω_t the temporal frequency. $\mathcal{F}\{\cdot\}$ denotes the Fourier transform, and δ denotes the Dirac delta function. Equation 2 states that all the energy associated with a translating pattern lies on a plane through the origin in the space of frequencies. Thus, we obtain the following constraint on the image velocity (Watson and Ahumada, 1985; Adelson and Bergen, 1985)

$$u\omega_x + v\omega_y = -\omega_t. \tag{3}$$

The left hand side of equation (3) denotes the dot product of the motion vector $\vec{v} = (u, v)$ and the spatial frequency vector $\vec{\omega_s} = (\omega_x, \omega_y)$. Thus, given a frequency triplet $(\omega_x, \omega_y, \omega_t)$, this equation provides a constraint on the motion component parallel to $\vec{\omega_s}$. To obtain the 2D motion vector $\vec{v}$, we need at least two measurements along different spatial frequency orientations.

To simplify the analysis, in the following sections we consider only images with bar-like structures parallel to the vertical dimension and the motion component perpendicular to the bars. Thus, let us consider a two-dimensional case of $I(x, t)$, that is a signal $I(x)$ which is shifted. Equation (1) then simplifies to

$$I(x, t) = I(x - ut, 0), \tag{4}$$

and the image motion constraint amounts to

$$u\omega_x = -\omega_t, \tag{5}$$

defining a line in the two-dimensional frequency space. The velocity u can be found from the ratio of the temporal and spatial frequency, i.e. as

$$u = -\frac{\omega_t}{\omega_x}. \tag{6}$$

Figure 3b, c illustrate the spatio-temporal signal for an image line in the Donguri pattern moving with velocity $u = 1$ and $u = -1$, respectively. Since the spatio-temporal signal $I(x, t)$ is obtained simply by shifting the signal

$I_0(x)$, it has the same structure in the spatial and temporal domain. Referring to Figures 3b and c, a spatial cross-section through $I(x,t)$ gives a shifted version of $I_0(x)$. A temporal cross-section gives a shifted, stretched and maybe reflected version of I_0. For motion to the left ($u = -1$) it is just the shifted image signal $I_0(t)$, and for motion to the right ($u = 1$), it is the shifted $I_0(-t)$ (i.e the reflection of $I_0(t)$). The amount of stretch encodes the motion. Thus, later when we analyze temporal filtering, instead of examining the temporal cross-section, we can look at the spatial cross-section.

3 The Filters

The spatio-temporal energy filters for extracting motion are separable in space and time. This just means that the filters can be created as the product between a spatial and a temporal filter. For the analysis this means that the spatio-temporal signal may first be convolved with the spatial filter and the result may then be convolved with the temporal filter.

The filters need to be localized in image space as well as in frequency space. We follow the common formulation of modeling a filter for detecting the local frequency ω_0, as a complex function

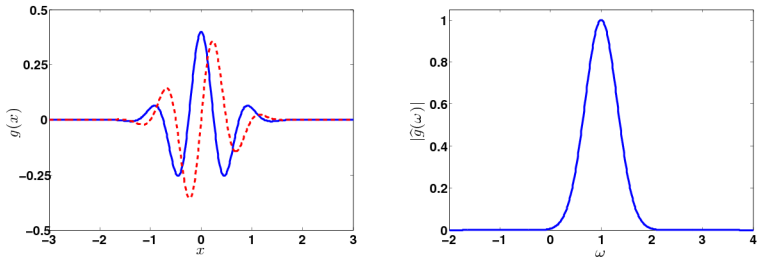
$$g(y) = p(y) \cdot exp(2\pi i \omega_0 y). \tag{7}$$

$exp(2\pi i \omega_0 y) = \cos(2\pi\omega_0 y) + i\sin(2\pi\omega_0 y)$, called the carrier function, is a complex sinusoidal for detecting the signal's component of frequency ω_0, and $p(x)$, called the envelope function, localizes the sinusoid in image space. The complex filter really consists of two filters in quadrature, the even cosine components and odd sine components. For example, in the spatial domain, the even component will respond maximally to bar-like signals and the odd component will respond maximally to edges. The magnitude of the output of the combined complex filter does not depend on whether the signal is even or odd, or any mixture thereof. As a result, complex motion filters (Adelson and Bergen, 1985; Watson and Ahumada, 1985) allow to extract motion independent of the phase of the signal, that is independent of the position of the signal within the receptive field at certain time, and independent of the sign of the contrast.

3.1 Modeling the spatial and temporal filters

We model the spatial filters as Gabor functions (see Figure 5a) with impulse response

$$G(x;\omega_x) = \frac{1}{\sqrt{2\pi}\sigma_x} exp(\frac{-x^2}{2\sigma_x^2}) \cdot exp(2\pi i \omega_x x), \tag{8}$$



(a) (b)

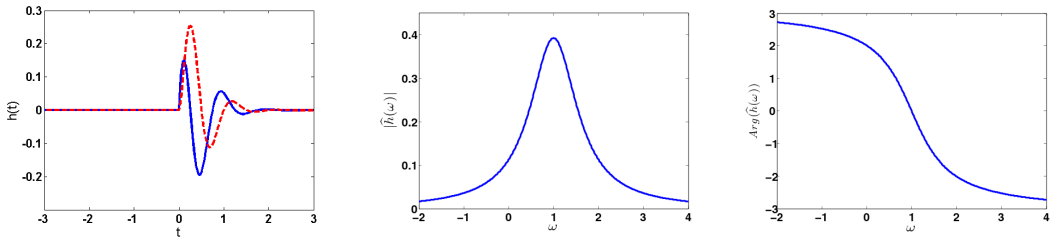
Fig. 5. Gabor filter of $\omega_x = 1$: (a) Impulse response. Blue denotes the real (even) part and red the imaginary (odd) part. (b) Amplitude spectrum.

where the envelope is a Gaussian. ω_x is the preferred frequency and σ_x determines the support of the filter, which from experience is chosen as $\sigma_x = \frac{1}{2\omega_x}$. The transfer function of the Gabor filter, which is obtained as its Fourier spectrum, amounts to $\hat{G}(\omega;\omega_x) = exp(-2\pi^2\sigma_x^2(\omega-\omega_x)^2)$, that is a Gaussian centered at ω_x and of standard deviation $\frac{1}{2\pi\sigma_x}$. The Gabor of frequency ω_x, thus extracts the signal's energy in a small frequency band around ω_x. Fig. 5b illustrates the amplitude of $\hat{G}$. Its phase is zero (i.e there is no imaginary part), because the envelope of the Gabor is symmetric around 0. In general, a filter which is symmetric around some other point (different from 0) has a phase linear in the frequency.

The temporal filter has an envelope described by a function with first order exponential decay. We use the formulation proposed in (Chen et al., 2001; Shi et al., 2004), which models the envelope as a Gamma probability density function of parameter $\Gamma(2)$, resulting in temporal filters $T(t)$ of the form

$$T(t;\omega_t) = \begin{cases} \frac{t}{\tau^2} exp(-\frac{t}{\tau}) \cdot exp(2\pi i\omega_t t) & \text{for} \quad t \geq 0 \\ 0 & \text{for} \quad t < 0 \end{cases} \tag{9}$$

where ω_t is the temporal frequency. τ, the decay velocity, is a time constant for the envelope, which we chose as $\frac{1}{4}\omega_t$ to make the wave of the temporal filter similar to the Gabor. Figure 6a illustrates the impulse response of this filter. Since the temporal filter extends from the past to the present, it actually estimates the frequency at a point earlier in time. This can also be seen from the spectrum of the filter. As can be observed from Figures 6b and c, the amplitude of $\hat{T}(t)$ is still a hat-type function, centered at ω_t similar to the Gabor. However, the phase of $\hat{T}(t)$ is non-zero, indicating a shift of the response in image domain. It is approximately linear for ω close to ω_t and deviates from linearity for values farther from ω_t. The next section explains, that this shift in the localization of the temporal frequencies in conjunction with imperfect localization of certain spatial frequencies leads to erroneous estimation of the



(a) (b) (c)

Fig. 6. Temporal filter of $\omega_t = 1$: (a) Impulse response. Blue denotes the real part, red the imaginary part (b) Amplitude spectrum. (c) Phase spectrum.

image velocity.

3.2 Definition of filtering

When analyzing image motion, we can think of the filtering as a spatial filtering followed by a temporal filtering. First, the image sequence $I(x,t)$ is filtered with a spatial Gabor, $G(x;\omega_x)$ to obtain the image sequence $\tilde{I}(x,t;\omega_x)$ as

$$\tilde{I}(x,t;\omega_x) = \int_{-\infty}^{\infty} I(y,t)G(x-y;\omega_x)dy.$$

At this stage it is assumed that the dominant spatial frequency of $\tilde{I}(x,t;\omega_x)$ at every point (x,t) is ω_x. Second, $\tilde{I}(x,t;\omega_x)$ is filtered with the temporal filter $T(t;\omega_t)$ to obtain $\tilde{\tilde{I}}(x,t;\omega_x,\omega_t)$ as

$$\tilde{\tilde{I}}(x,t;\omega_x,\omega_t) = \int_{-\infty}^{0} \tilde{I}(x,s;\omega_x)T(t-s;\omega_t)ds.$$

Here it is assumed that $|\tilde{\tilde{I}}(x,t;\omega_x,\omega_t)|^2$ returns the motion energy of I at image point (x,t) at frequencies (ω_x,ω_t).

Since the spatial and temporal filters are complex valued, the complete spatio-temporal filter can be imagined as four separable filters (the even and odd components of each, the spatial and temporal filter), whose outputs are summed according to the rules of complex numbers to arrive at the motion energy.

3.3 Image motion estimation

In the following we will analyze motion estimation as a function of spatial frequency. In this analysis we compute the image motion of a patch in two

stages. First, we estimate at every point the (best) velocity. Second, we compute the velocity of the patch as the weighted average of point-wise velocity estimates.

Specifically, given a spatial frequency ω_x, the signal $I(x,t)$ is filtered with the spatial Gabor to obtain $\tilde{I}(x,t;\omega_x)$. Then (at some time t, and it does not matter which) at every image point x, we find the dominant temporal frequency ω_{t_0}. In other words, we search for the temporal filter $T(t,\omega_{t_0})$, which among all temporal filters $T(t,\omega_t)$ of different frequencies ω_t returns the maximum energy. (Ideally according to the motion constraint (eq. 2), only one temporal frequency filter should return non-zero energy.) This way we find at every point x a velocity estimate $\hat{u}(x;\omega_x) = -\frac{\omega_{t_0}}{\omega_x}$ and energy $|\tilde{\tilde{I}}(x,t;\omega_x,\omega_{t_0})|^2$.

The motion of a patch then is found as the average of energy weighted velocity measurements:

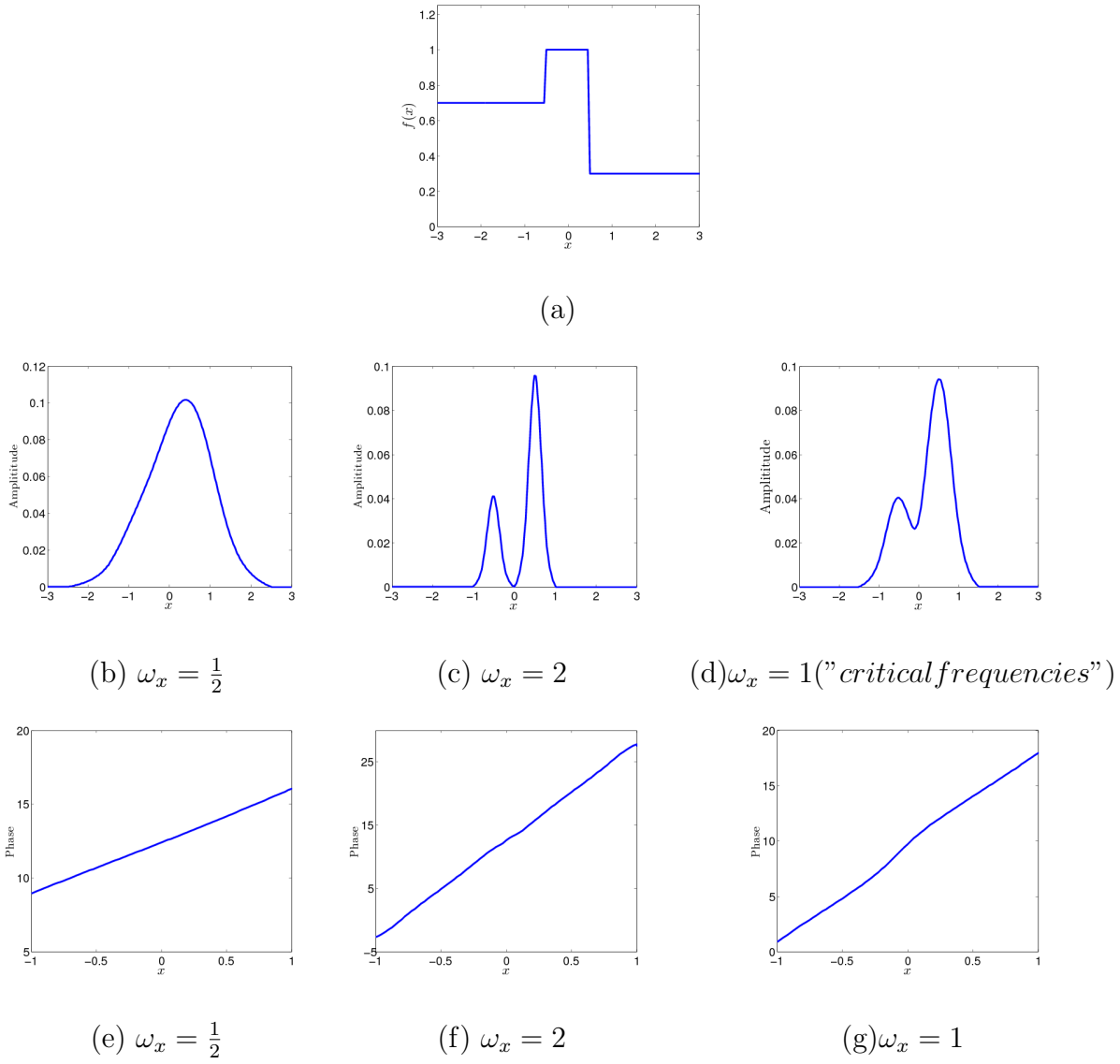
$$\hat{u}(\omega_x) = \frac{\sum_x \hat{u}(x;\omega_x)|\tilde{\tilde{I}}(x,t;\omega_x,\omega_{t_0})|^2}{\sum_x |\tilde{\tilde{I}}(x,t;\omega_x,\omega_{t_0})|^2}. \tag{10}$$

4 The Effect of Filtering on Donguri

Figure 7 illustrates the effect of spatial filtering on Donguri. At frequencies ω_x larger than the width of the bar, the Gabor filter detects the two edges at the left and right of the bar (Fig. 7b). At frequencies significantly smaller than the width of the bar, the Gabor detects the bar (Fig. 7c). The amplitude of the response thus has either one or two well separated peaks. However, for frequencies of ω_x close to the width of the bar, there is something in between one and two responses. The amplitude function becomes asymmetric with two merging peaks, a larger on the right and a smaller on the left (Fig. 7d). Let us call these frequencies the "critical frequencies".

The corresponding phase responses (Fig. 7e-g) are (nearly) linear for $\omega = \frac{1}{2}$ and $\omega = 2$, but the response significantly deviates from linearity for the critical frequencies, which is an indicator for poorly estimated frequencies. As is well known from the uncertainty principle, there is a limit on the accuracy of localization in image and frequency domain. The Gabor (which is the filter with best localization in joint image and frequency space) cannot guarantee perfect localization of the signal. Because of the "hat" profile of its Gaussian envelope, the filtered signal $\tilde{I}(x,t;\omega_x)$ will not always have local dominant frequency ω_x. It could be some value close to ω_x. To evaluate the local frequency of $\tilde{I}(x,t;\omega_x)$ at every point x, we filter this signal with a range of Gabor filters and find the frequency resulting in maximum energy.

Referring to Fig. 8 for $\omega = \frac{1}{2}$ and 2 the dominant local frequencies are esti-



(a)

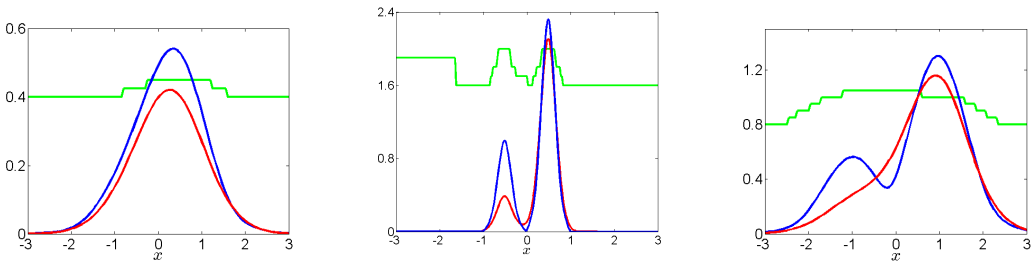
(b) $\omega_x = \frac{1}{2}$ (c) $\omega_x = 2$ (d)$\omega_x = 1$("$critical frequencies$")

(e) $\omega_x = \frac{1}{2}$ (f) $\omega_x = 2$ (g)$\omega_x = 1$

Fig. 7. (a) Bar in Donguri pattern. (b-d) Amplitude of bar filtered with Gabor of different frequencies ω_x. (e-g) Corresponding phase response.

mated quite well at the locations of high energy. Most important, however, the local frequency estimates are symmetric. There are either one or two symmetric peaks. In comparison, for the critical frequencies, the estimates are not symmetric. There is higher frequency to the left of the main peak than there is to the right of this peak. Using this observation, we can now state what constitutes an asymmetric signal:

A signal is asymmetric at certain spatial frequency, if the spatially (Gabor) filtered signal has an asymmetric amplitude response and an asymmetric local frequency response.



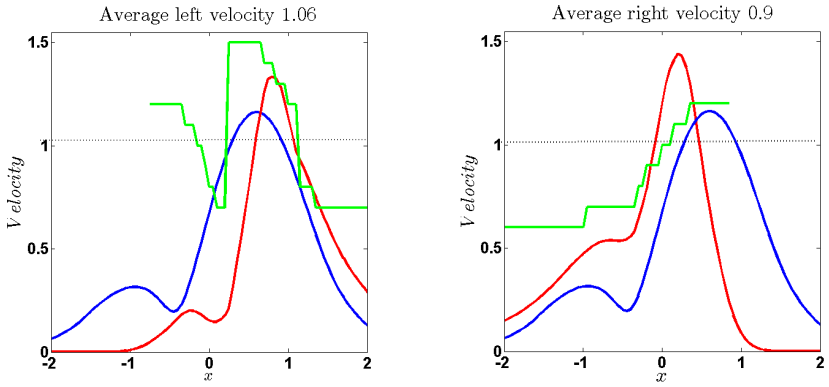
$\omega_x = \frac{1}{2}$ $\omega_x = 2$ $\omega_x = 1$

Fig. 8. Local frequency of $\tilde{I}$. Blue denotes the amplitude of $\tilde{I}$, green denotes the dominant local frequency, and red denotes the corresponding amplitude.

Remember, in estimating motion we assume that the spatial frequency of $\tilde{I}(x,t;\omega_x)$ is ω_x. When asymmetric temporal filters are used, the mislocalization of spatial frequency has an effect on the estimated image motion. Because of the asymmetry, left and right motion are not estimated the same. Figure 9 shows that for the critical frequencies, motion to the left ($u=-1$) leads to larger velocity estimates than motion to the right ($u=1$).

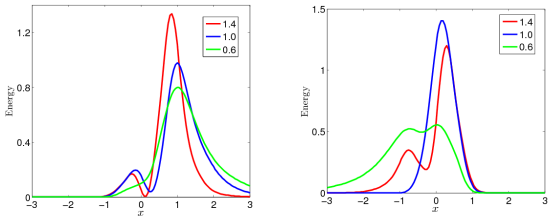
Let us give an intuition for this mis-estimation. The signal $\tilde{I}(x,y)$ (Fig. 7d), whose amplitude has a larger lobe on the right and smaller lobe on the left is convolved with the temporal filter (Fig. 6a). Recall, that convolution, is the inner product of the signal with the reflection of the filter upon zero. Thus, after reflection, for the case of left motion, the filter extends to the left of zero. For the case of right motion the temporal filter extends to the right of zero. Referring to Fig. 9, when convolving the two asymmetric signals, there is overestimation of image motion at the left peak for left motion, and underestimation of image motion at the right peak for right motion. This is also demonstrated in Fig. 10, which shows the estimated energy for three different temporal frequencies. As a result, the average velocity is larger for left motion than for right motion. The main reason is interaction of the regions under the two peaks with each other during temporal filtering, which however is different for the two directions.

For lower spatial frequencies (Fig. 8b) there is only one peak, and for higher spatial frequencies (Fig. 8c) the peaks are well separated and do not interact. Thus, in both cases there is no significant difference between left and right motion.



left motion right motion

Fig. 9. Velocity estimation at critical frequencies. Blue denotes the amplitude of $\tilde{I}$. Green denotes the local estimated velocity and red denotes the amplitude of $\tilde{\tilde{I}}$. (Note, because of the asymmetry of the temporal filter the maximum value for $\tilde{\tilde{I}}$ is found to the right of the stronger edge for left motion and to the left of the stronger edge for right motion). The estimated average velocity (estimated using eq. 10) is larger for left than for right motion.



left motion right motion

Fig. 10. Amplitude of $\tilde{\tilde{I}}$ for temporal frequencies $\omega_t = 0.6, 1.0$ and 1.4.

5 Experimental Evaluation

5.1 Donguri and Rotating Snake

The following figures show the estimated image velocity as a function of spatial frequency. The estimates were obtained by simulations as described in 3.3. That is, the motion of a pattern element is computed as the energy weighted average of the local velocity at every image point (eq. 10). Fig. 11a shows the estimated velocity for left and right motion for a large range of frequencies, demonstrating that significant differences occur in a small range around the frequency $\omega = 1$. Fig. 11b zooms in on a neighborhood around the critical frequencies, but shows only the difference in velocity between estimated left and right motion. Let us clarify, higher frequencies (of the Gabor filter) in

our plot correspond to higher resolution of the perceived image, that is filters located closer to the fovea.

Referring to Figure 11a, the estimates fluctuate in the neighborhood of the critical frequencies. There is an overestimate for left motion and an underestimate for right motion at $\omega = 1$. Both velocities are overestimated for a bit larger ω (1.25), and both are underestimated for a bit smaller ω (0.75), but at these frequencies their differences are not significant. To test the validity of the approach, we experimented by varying the parameters in the motion estimation. In particular, we varied the range of possible temporal frequencies (with the smallest range [0.6..1.4] and the largest unlimited), the size of the spatial and temporal filters, and the weighting of the local velocity estimates. (Besides using the energy we used the absolute value of the filter response and its cube.) We found that for some parameter settings, both left and right motion were underestimated at $\omega = 1$. However, for all settings, there was a significant difference at the critical frequency, with the left motion being larger than the right. Based on these experiments, we state the gist of our finding as: Estimated left motion is larger than estimated right motion for the critical frequencies.

Next, consider the pattern Rotating Snake (Fig. 2) and take three cross sections through one of its units to obtain three qualitatively different profiles. Referring to Figure 12, the first cross-section is at the center, providing a profile just like Donguri's, the second cross-section is in the upper half, where all intensity regions have equal width, and the third is close to the top of the unit, where the intermediate intensity regions (yellow and blue) become narrow bars and the black and white regions have large extent. We refer to these profiles as "Donguri", "Bars", and "Steps", respectively. Linear arrangements of the corresponding monochromatic signals are shown in Fig. 12b-c. Figure 13 plots the difference in estimated image velocity between left and right motion for the three functions. The simulations show that for all three signals there is a range of frequencies for which left motion is significantly larger than right motion. The difference is significantly larger in Donguri than Bars, and is larger in Bars than Steps.

5.2 Nulling experiment

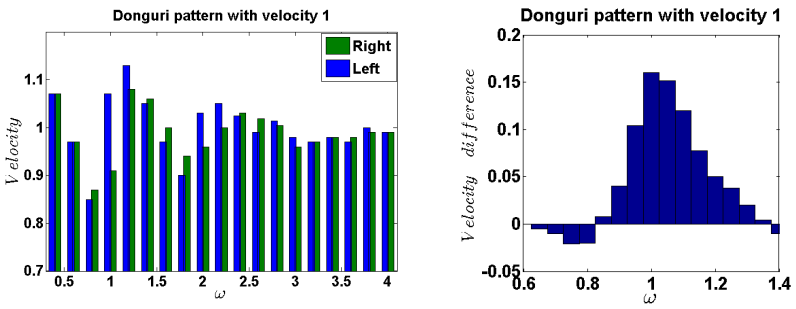
The strength of the illusory perception varies significantly between observers. The relative strength of the perceived motion in different signals, however, can be used to evaluate the model. We quantitatively compared the perception of the three signals above by nulling the illusory motion with opposing real motion, similar as in Murakami et al. (2006). 9 naive subjects participated in the experiment.

The signals were arranged on three concentric rings, with the middle ring three times the width of the inner and outer rings. Each ring consisted of 40 signal elements, and the flanking rings were phase shifted with respect to the middle ring by a quarter of the element, (Figure 15). The intensity values of the four regions were 0.3, 1, 0.7, 0. The width of the bar was 1 unit and the other regions were 3 units in Donguri and Stairs, and all regions were 2 units in Bars (as shown in Fig. 12). In addition to these three signals, we also tested a Donguri signal of reduced contrast with the intensities 3/8, 1, 5/8, 0. Simulations for this signal (Fig. 14) show that although the range of frequencies where left and right motion are significantly different is smaller than in the original Donguri, the predicted value at the maximum is nearly the same, actually slightly larger.

Observers were sitting about 45 centimeters in front of the screen. At this distance the rings covered 14° of visual field with the width of the three rings covering 1.7°. For each signal two patterns were created, one with the intensity regions in the order shown above inducing counterclockwise motion, and one obtained as the mirror reflection of the former, inducing clockwise motion.

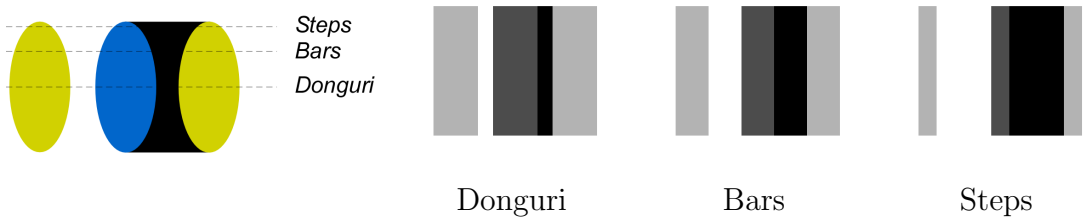
An interface was created that allowed to play videos showing these patterns rotating slowly both clockwise and counter-clockwise. The speed of motion could be set in the range of $0.06° - 0.6°/s$ by the step of 0.06°, where $1°/s$ corresponds to one degree of polar angle per second.

The speed of motion that gave the subjective stationary percept was found with the Staircase method. Observers were presented with videos of the drifting patterns increasing and decreasing in speed and asked to make a judgement on whether they experienced clockwise, counterclockwise or no motion. Figure 16 displays our measurements. The speed nulling illusory motion in Donguri was in the range of 0.12 - 0.36. All participant perceived Donguri the strongest, and Snake stronger than Stairs. None of the subjects measured a significant difference between the original Donguri and Donguri with decreased contrast, and between clockwise and counterclockwise stimuli. Fig. 17 compares the predictions to the mean of measurements. The values shown are the ratio of the nulling motion in Donguri to Snake, Bars and Reduced Donguri, and were obtained as the average over clockwise and counterclockwise patterns over all subjects. The predicted motions were found as the estimated maximum difference between left and right motion over all frequencies. The figure demonstrates that our model predicts the observations very well.



(a) (b)

Fig. 11. (a) Velocity estimation for Donguri. (b) Difference in velocity estimates between left and right motion.



Donguri Bars Steps

Fig. 12. Three different illusory intensity signals in Snake.

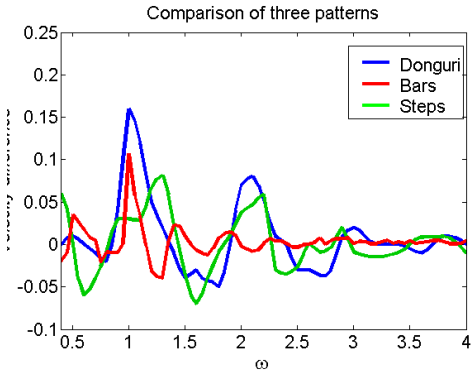


Fig. 13. Difference in estimated velocity between left and right motion.

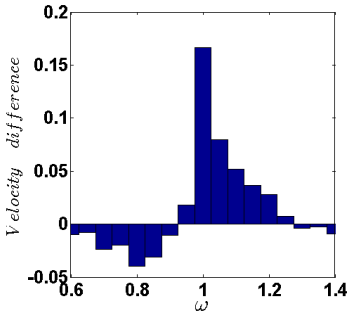


Fig. 14. Difference in estimated velocity for Reduced Donguri.

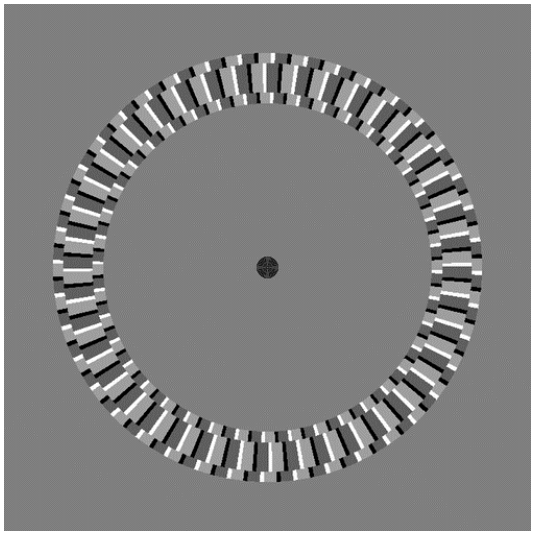


Fig. 15. Experimental stimulus (Donguri)

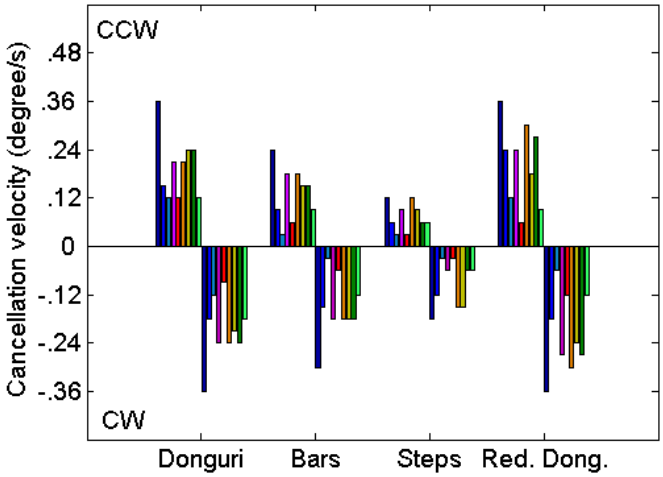


Fig. 16. Cancellation velocity in 9 subjects for clockwise and counterclockwise illusory motion. Each color corresponds to one subject

5.3 Peripheral Drift and Central Drift

It is generally considered that the illusory motion effect was first observed in the Peripheral Drift illusion (Fraser and K.J.Wilcox, 1979) (see Fig. 18). The intensity profile in this pattern is a sawtooth function. Such a function would not give rise to erroneously estimated motion according to our model. However, luminance recorded at the neural level is usually modeled as a non-linear function of the actual intensity of the image. Following (Bacus and Oruç, 1995), we consider two factors in our model: first, luminance adaption, a logarithmic function modeling the relationship between recorded luminance and physical intensity (changes at higher intensity values are recorded with a

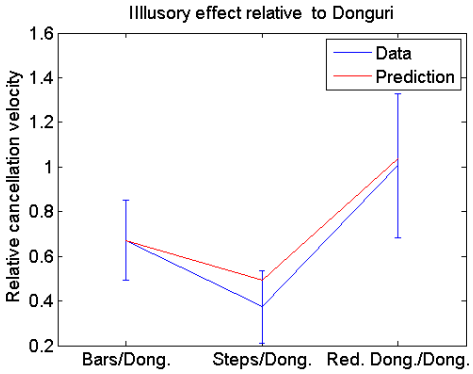


Fig. 17. Comparison of prediction to experimentally obtained velocity ratios of Bars : Donguri, Steps : Donguri, and Reduced Donguri : Donguri.

smaller value than changes at lower intensity values); second contrast adaptation, a sigmoid function modeling greater sensitivity to the middle range than the high and low ranges of intensities (see Fig. 19).

Kitaoka and Ashida (2004) created a series of patterns, which they call central drift illusions, as they are perceived in central as well as peripheral vision. For examples, see Sakura and Cendri in Fig. 21. These patterns contain elements (the petal and ovals) with (close to) linear intensity profiles, but in comparison to the Peripheral Drift illusion, the individual elements are separated by uniform background. This separation increases the illusory effect.

Applying our luminance model to the actual intensities, we obtain the luminance profiles shown in Figure 20. Our model's predicted velocity differences between left and right motion for Sakura are shown in Figure 22. The petals in the model are four times the width of the bar in Donguri. Thus, the critical frequencies of $\omega = 1$ in Fig.22 corresponds to the period of the sinusoid being a $\frac{1}{4}$ of the petal size. Figure 22b compares the velocity differences in Peripheral Drift, Sakura, and Cendri for a small range around the critical frequencies. Our model predicts perceived illusory motion in the patterns, with the effect being stronger in Sakura and Cendri than in Peripheral drift

A similar signal, which Kitaoka (2006) calls Type I , consisting of either white to medium gray elements on dark background, or medium to dark gray elements on white background, causes illusory motion in the periphery (Fig. 23a). The corresponding profiles may be considered smooth versions of the Donguri-profile. A simulation of the motion estimate, considering our luminance model predicts the perceived motion (Fig. 23c) Figure 24 shows one of Kitaoka's patterns from this class, in which the two elements are combined for an even stronger effect.

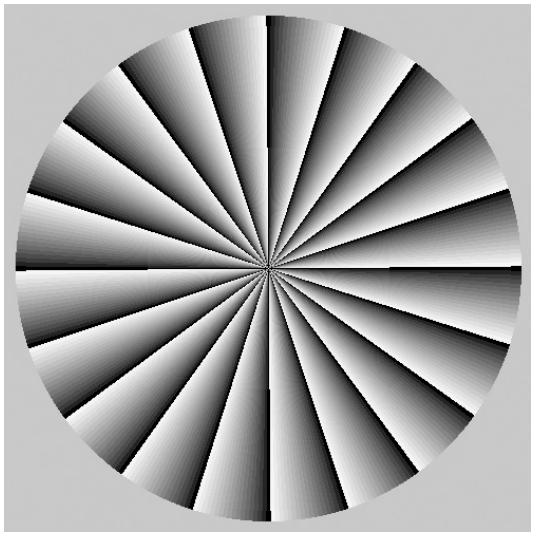
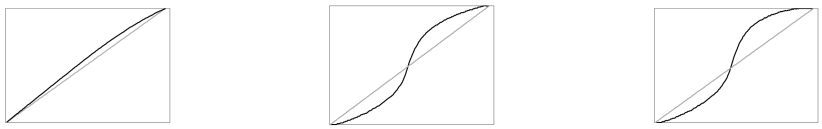
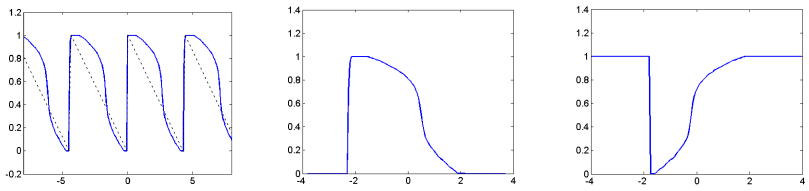


Fig. 18. The peripheral drift illusion (Fraser and K.J.Wilcox, 1979). In peripheral vision, the circle appears to rotate slowly in clockwise direction.



Luminance adaptation Contrast adaptation Compound adaptation

Fig. 19. Model of neural luminance function.



Drift Sakura W. Cendri

Fig. 20. Modeled luminance profile in peripheral and central drift patterns.

6 2D motion

When estimating 2D image motion we still have a two-stage computational process. In the first stage causal filters estimate point-wise erroneous motion in the direction perpendicular to the spatial filter orientation (the 1D motion component, also called normal flow). In the second stage normal flow estimates in different directions within spatial local neighborhoods are combined

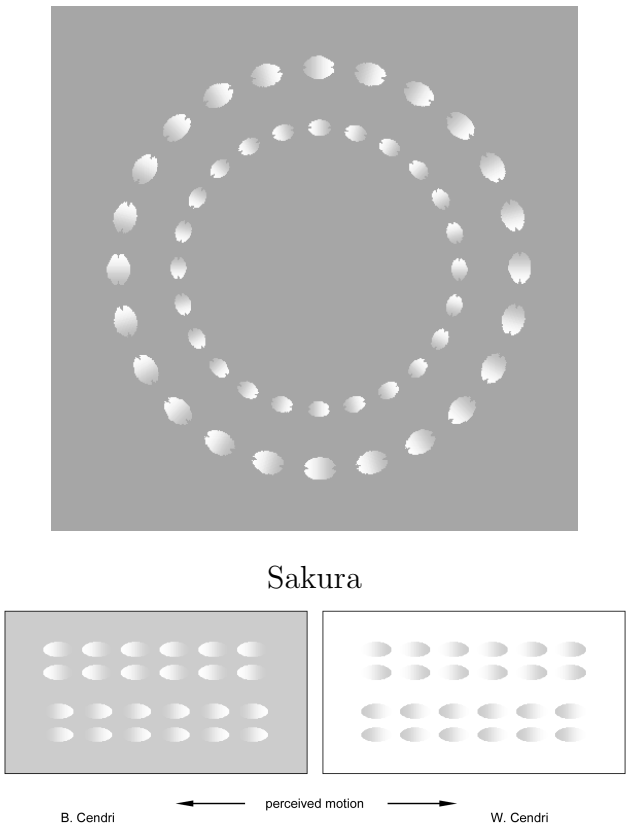
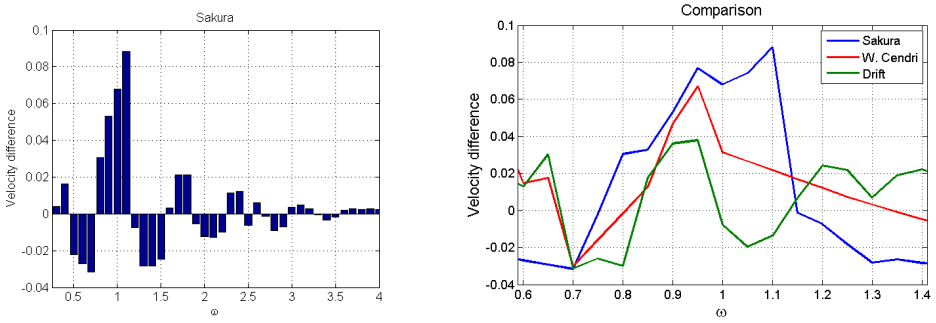


Fig. 21. Sakura in gray: When fixating on the center, the outer petals appear to move slowly clockwise and the inner petals move counterclockwise. B. Cendri (same as Sakura): the ovals on gray background move from gray to white. W. Cendri: the ovals on white background move from white to gray.



(a) (b)

Fig. 22. (a) Difference in estimated velocity between left and right motion in Sakura (b) Comparison of motion estimation between peripheral and central drift illusions.

to estimate the 2D image motion vector.

In the following simulations, we obtained the flow fields in response to horizontal and vertical rigid motions as follows: We first computed at the critical spatial frequency of the pattern, at every image point for four orientations (4

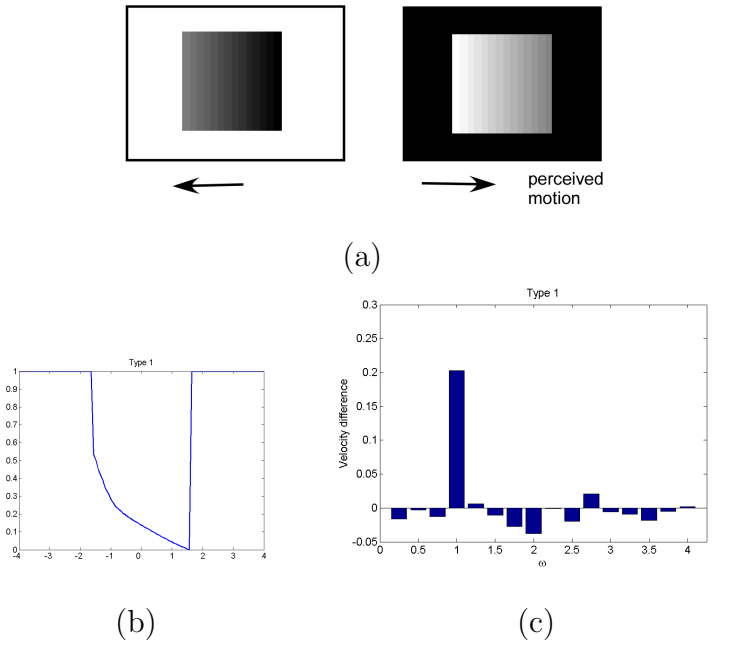


Fig. 23. (a) Gray shaded elements on bright and dark background. Dark elements on white move from dark to light. Light elements on black move from light to dark (This is opposite to the peripheral drift illusion!) (b) Luminance profile for the dark element on white background. Note, the element is chosen of size 3. (c) Difference in estimated velocity between left and right motion.

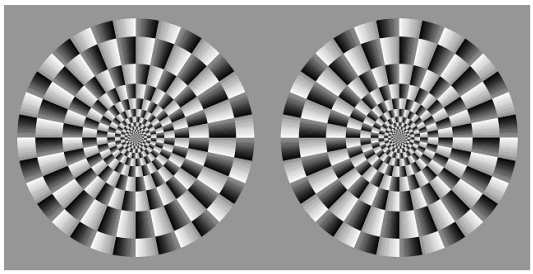


Fig. 24. Type I illusion (Kitaoka, 2006)

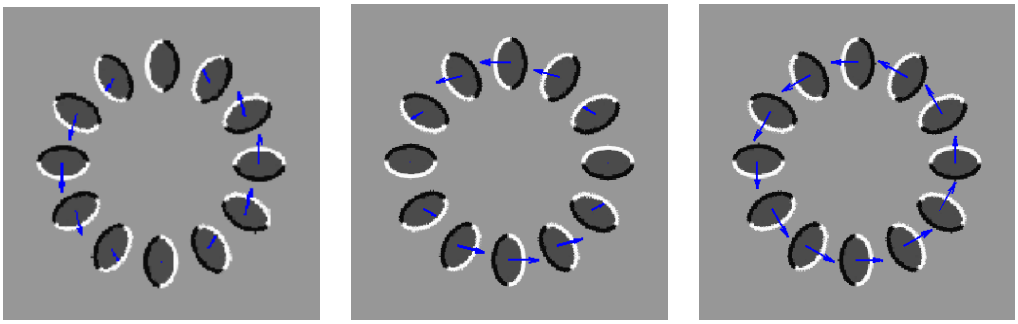
Gabors) the corresponding temporal frequency using causal filters, to arrive at n equations of the form

$$\omega_{x_i} + \omega_{y_i} v = -\omega_{t_i} \quad \text{with} \quad i = 1 \ldots n. \tag{11}$$

Then we obtained the flow (u, v) of the pattern element by solving the overdetermined system of n equations using least squares estimation, i.e. as

$$\begin{pmatrix} x \\ y \end{pmatrix} = (W_x^T W_x)^{-1} W_x W_t \tag{12}$$

with W_x the $n \times 2$ matrix $(\omega_{xi}, \omega_{y_i})$ and W_t the n-dimensional vector (ω_{ti}).



(a) (b) (c)

Fig. 25. (a) Average flow for Donguri at critical frequencies for (a) horizontal (left and right) (b) vertical (up and down) (c) combined horizontal and vertical motion.

Figure 25a shows the average flow vectors resulting from left and right motion on the Donguri pattern. In other words, the vectors shown are the vector sum from a left and a right movement. Figure 25b shows the average flow for up and down motion, and Figure 25c shows the average flow for all four motions.

7 Conclusions

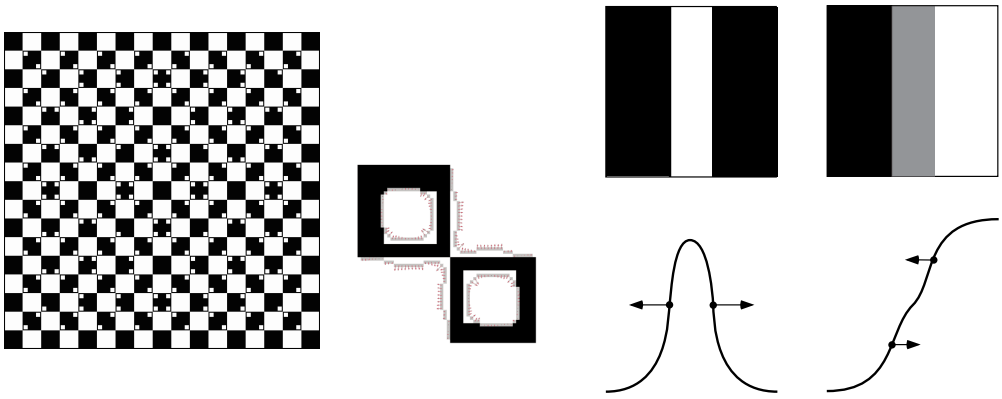
7.1 Summary

Temporal image motion filtering is causal. That is, the filters use data from the past, but do not use data from the future. In this paper we showed that this asymmetry in the filters leads to erroneous estimation of image motion for asymmetric signals at certain scale. This is simply because of the universal uncertainty in estimating signals. We demonstrated the mis-estimation using simulations. Then we tested our model quantitatively using different signals with bar-like structures and found that it very well predicts the illusory motion perception. Based on these findings, we hypothesize that this erroneous estimation is the cause for illusory perception of motion in static patterns with repeated asymmetric pattern elements.

7.2 Relationship to geometric optical illusions

The concept of smoothing at certain scale as an explanation for optical illusions is not new to the literature. Morgan and Moulden (1986); Morgan and Casco (1990) have proposed that bandpass filtering (that is edge detection by computing derivatives on a smoothed image) is the cause of a number of static geometric optical illusions. For an example see Fig.26a. The illusory elements

are bars. As discussed in Fermüller and Malm (2004), if we smooth a bar with a Gaussian of σ large enough to effect both edges of the bar but not large enough for the two edges to merge, the location of the edges changes, as illustrated in Fig. 26c. For a bright bar in a dark region (or a dark bar in a bright region) the two edges drift apart. For a bar of medium brightness next to a bright and a dark region the two edges move toward each other. The latter case corresponds to the Donguri profile. The σ in the Gaussian is $\sim \frac{1}{2}$, which is the same as the σ in the Gabor of the "critical frequencies". Thus, at the "critical frequencies" the interaction of the two edges causes a change in the location of the edges (defined as the extrema in the first order derivatives or zero-crossings in second order derivatives). At the same time, as we showed in this paper, local frequencies are poorly estimated, which has an effect if image sequences are filtered asymmetrically in temporal domain.



(a) (b) (c)

Fig. 26. (a) Illusory pattern "waves"– a perfect checkerboard pattern with superimposed squares – causes the perception of wavy lines (Kitaoka, 1998) (b) Demonstration of the movement of edges under smoothing for a small part of the pattern. (c) A schematic description of the behavior of edge movement in scale space. The first row shows the intensity functions of the two different bars, and the second row shows the profiles of the (smoothed) functions with the dots denoting the location of edges, which either drift apart or get closer.

References

Adelson, E. H., Bergen, J. R., 1985. Spatiotemporal energy models for the perception of motion. Journal of the Optical Society of America A 2, 284–299.

Albrecht, D., Geisler, W., 1991. Motion selectivity and contrast response function of simple cells in the visual cortex. Visual Neuroscience 7, 531–546.

Albrecht, D., Geisler, W., Frazor, R., Crane, A., 2002. Visual cortex neurons of monkeys and cats: Temporal dynamics of the contrast response function. Journal of Neurophysiology 88, 888–913.

Anstis, S., 1970. Phi movement as a subtraction process. Vision Research 10, 1411–1430.

Ashida, H., Kitaoka, A., 2003. A gradient-based model of the peripheral drift illusion. In: Proc. ECVP. Paris.

Bacus, B. T., Oruç, I., 1995. Illusory motion from change over time in the response to contrast and luminance. Journal of Vision 5 (11), 1055–1069.

Burr, D., Morrone, M., 1993. Impulse response functions for chromatic and achromatic stimuli. JOSAA 10, 1706.

Chen, Y., Wang, Y., Qian, N., 2001. Modeling v1 disparity tuning to time-varying stimuli. Journal of Neurophysiology 86, 143–155.

Conway, B., Kitaoka, A., Yazdanbakhsh, A., Pack, C., Livingstone, M., 2005. Neural basis for a powerful static motion illusion. Journal of Neuroscience 25, 5651–5656.

Eizenman, M., Hallett, P., Frecker, R., 1985. Power spectra for ocular drift and tremor. Vision Research 25, 1635–1640.

Fermüller, C., Malm, H., 2004. Uncertainty in visual processes predicts geometrical optical illusions. Vision Research 44, 727–749.

Fraser, A., K.J.Wilcox, 1979. Perception of illusory movement. Nature 281, 565–566.

J. Faubert, J., Herbert, A. M., 1999. The peripheral drift illusion: A motion illusion in the visual periphery. Perception 28, 617–621.

Kitaoka, A., 1998. `http://www.ritsumei.ac.jp/~akitaoka/cushione.html`.

Kitaoka, A., 2003. `http://www.psy.ritsumei.ac.jp/~akitaoka/rotsnakee.html`.

Kitaoka, A., 2006. The effect of color on the optimized Fraser-Wilcox illusion. Gold prize at the 9th L'ORAL Art and Science of Color Prize.

Kitaoka, A., Ashida, H., 2003. Phenomenal characteristics of the peripheral drift illusion. Vision 15, 261–262.

Kitaoka, A., Ashida, H., January 2004. A new anomalous motion illusion: the "central drift illusion". In: Winter meeting of the Vision Society of Japan.

Morgan, M. J., Casco, C., 1990. Spatial filtering and spatial primitives in early vision: An explanation of the Zöllner-Judd class of geometrical illusions. Proc. Royal Society, London B 242, 1–10.

Morgan, M. J., Moulden, B., 1986. The Münsterberg figure and twisted cords. Vision Research 26 (11), 1793–1800.

Murakami, I., 2004. Correlations between fixation stability and visual motion sensitivity. Vision Research 44, 251–261.

Murakami, I., Cavanagh, P., 1998. A jitter after-effect reveals motionbased stabilization of vision. Nature 395, 798–801.

Murakami, I., Kitaoka, A., Ashida, H., 2006. A positive correlation between fixation instability and the strength of illusory motion in a static display. Vision Research 46, 2421–2431.

Ross, J., Morrone, M., Goldberg, M., Burr, D., 2001. Changes in visual perception at the time of saccades. Trends in Neurosciences 24 (2), 113–121.

Shi, B. E., Tsang, E. K. C., Au, P. S. P., 2004. An on-off temporal filter circuit for visual motion analysis. In: ISCAS (3). pp. 85–88.

Watson, A. B., Ahumada, A. J., 1985. Model of human visual motion sensing. Journal of the Optical Society of America 2, 322–342.