Motion Illusions in Man and Machine

Cornelia Fermüller

Institute for Advanced Computer Studies
University of Maryland
College Park, MD 20742, U.S.A.

Abstract

The computational problem of motion perception involves early processes of computing image motion from the retinal images, and later processes of interpreting the image motion in terms of the 3D motion of the observer and the objects in the scene. At the level of mathematical abstraction, computing image motion amounts to an estimation problem and can be analyzed using the tools of statistics and signal processing. As shown in this chapter, there are intrinsic limitations to the estimation processes that make it impossible to derive veridical estimates for all images. We propose that this is the main reason for many optical illusions of motion perception on static image patterns. Image motion is estimated erroneously, and as a result higher level processes arrive at an interpretation of erroneous three-dimensional motion and moving scene. Specifically, we discuss two limitations. First, because of noise in the image data, there is statistical bias in the estimation of image motion leading to consistent erroneous estimates. The effect is largest in textured regions of one dominant gradient direction and can account for motion illusions in static patterns of line drawings, such as the *Ouchi illusion*. Second, because biological motion is real-time, the filters for estimating image motion are symmetric in space but asymmetric (causal) in time. In other words, only the past but not the future is used to estimate the change in the temporal domain. This leads to errors in image motion estimation in locally asymmetric intensity signals of certain spatial frequencies and can explain the effect in patterns of asymmetric intensity profiles, such as the *Snake illusion*. Since these limitations are not an artifact of the hardware, but are inherent to the computations, they will affect any system, and thus create illusions in man and machine.

Key words: illusory motion, computational modeling, image motion estimation, spatio-temporal filtering statistical bias

1 Introduction: Theories of Illusions

Optical illusions have been studied in visual psychology since its early history, and the formation of theories followed closely. Since then there has been an enduring interest in the subject. Theorists of perception have looked at illusions as test instruments for theory, an effort that originated from the founders of the Gestalt school. An important strategy in finding out how correct perception operates is

to observe situations in which mis-perception occurs. Any theory, to be influential, must be consistent with the facts of correct perception but also must be capable of predicting the failures of the perceptual system. While in the past, only psychologists have studied illusions, in recent decades they have been joined by scientists of other mind-related fields such as neurology, physiology, philosophy, and the computational sciences, examining the problem from different viewpoints with the use of different tools (Gillam, 1998; Palmer, 1999).

It is clear that vision cannot be veridical, or as Palmer (1999, chap. 1) puts it, vision cannot give us "a clear window into reality." Visual perception is very complex as it includes a large set of processes that start at the formation of images on our eyes and provide us with an interpretation of the objects and events in the scene. For the purpose of discussion, we distinguish three kinds of processes: processes of image formation, processes of building geometric models of the spatio-temporal scene, and processes of interpreting the scene in terms of the concepts we know. According to this division, we can classify illusions of vision into *physiological illusions* due to the image formation or very early representation of images, *optical illusions in the classic sense*, which are about perceiving surfaces and motions different from the real objects, and *cognitive illusions* resulting from an ambiguous interpretation.

It does not surprise us that there are limitations to the early processes. Our eyes can be likened to an optical apparatus that collects the light emitted from surfaces in the world, and like any measuring device, the eye has its limitations. It is limited by the spatial and temporal resolution of the photo-receptors and their adaptive sensitivity to light. These limitations can account for many of the physiological illusions; for example afterimages, the apparent motion effect, or the illusion of a fast rotating object appearing to rotate in opposite direction of its true movement (the wagon-wheel effect). It is also no surprise that our interpretation of scenes is not unique. Cognitive interpretation requires high level knowledge and memory. This knowledge is acquired from our experiences, and thus can vary between individuals. Cognitive illusions may help us understand the attention mechanisms involved in complex interpretation and how our mind organizes high-level visual information. Examples of such illusions are ambiguous figures, such as the old-young woman illusion, or paradox illusions such as the impossible staircase.

The illusions most studied by vision scientists, are the optical illusions in the classic sense, which are due to the processes of making models of the scene, or as computational scientists say, the processes of "reconstructing the scene." Illusions in this category include the geometric-optic illusions, illusions of lightness and color, visual completion illusions, and illusions of motion. One not familiar with the study of vision may think that reconstruction is simple, but clearly it is not. The problem often is summarized as "the vision system has to solve the inverse optics problem." The images formed on our retina are projections of the 3d world on 2d imaging surfaces, and it is up to the vision processes to reconstruct the third dimension. The light incident on the retina depends on many factors, such as the different light sources, the properties of the reflecting surfaces and inter-reflections between objects, and it is up to the vision processes to separate the different lighting components. The image intensity pattern in a single image may not carry enough information to divide the image into different surfaces and objects, and it is up to the vision system to use the best information to do so. In these processes of recovering the surfaces in the world and the movements of the scene from the two-dimensional light patterns, things can go wrong and misperception occurs, giving us a window into the mechanisms underlying the reconstruction.

Most works on computational modeling have been influenced by the computational theory of Marr (1982). This theory describes vision as information processing, which must be understood at

three levels: at the computational level (What does the system do and why does it do it?), at the algorithmic or representational level (What are the processes and representations?) and at the physical level (What is the physical realization, i.e. the hardware, or the neural structure of biological systems?) Marr suggests that what the system computes is three-dimensional (3d) descriptions of the world from two dimensional (2d) images, and this is algorithmically implemented in a sequence of stages. First local features, such as line features, feature points, and local image motion are extracted, these are linked to obtain contours and connected regions, then the shape of the surfaces as seen from a given viewpoint is derived from cues such as motion, texture or shading, and finally the scene models from multiple views are combined into a single 3d model.

Most computational theories of optical illusions seek explanations either at the algorithmic level or the physical level. For example, theories of lightness illusions and geometric optical illusions predominantly refer to the representations, usually the quantities at the first stage of the Marr representation, such as edges, lines, junction corners, and region, and the mechanisms of computing them (e.g. Grossberg and Todorovic, 1988; Morrone and Burr, 1988; Neumann and Mingolla, 2003), and many theories of motion illusions refer to the hardware mechanisms implementing the two-dimensional image motion (e.g. Mingolla, 2003; Conway et al., 2005). While these theories seek specific mechanisms, which are hypothesized to exist in the vision system, the theories proposed here are at a higher level of abstraction. We argue the reason for the malfunctioning of image motion estimation is to be found at the level of abstract computational modeling, no matter how the computation is algorithmically or physically implemented.

At the computational level, many theories of optical illusions do not consider the process of obtaining world models from images. Since many optical illusions are patterns on a two-dimensional surface, it is often assumed that the goal of vision is to create an accurate image of these 2D dimensional patterns. This applies for most classical theories. A notable exception is Gregory's theory of geometric optical illusions, which supposes that 2d patterns are interpreted as flat projections of three-dimensional scenes. In the case of motion, reference to a 3d interpretation is imminent. It is clear that motion is a cue to space-time geometry. The goal of early motion interpretation is to build spatio-temporal models of the scene, and considering the 3d interpretation as the goal of motion processing, as explained next, can explain why independent motions are seen in static images.

2 Motion Perception

We can see the motion in the world because our vision has mechanisms to perceive changes in the image signal over time. It has been proposed that the human vision system has multiple mechanisms (Chubb and Sperling, 1988), some tuned to instantaneous local motion, some to regional motion, and some to large motion of segmented parts. The prominent mechanism, which allows us to see continuous motion, is based on computing the instantaneous changes of the incoming light. The vision system records the changes in the incoming light on the retina and then relates them in further processes to the physical changes in the world.

There are limitations to this motion mechanism, simply because of physical limitations, or "hardware" constraints. The motion mechanism is limited in temporal resolution. For this reason motion perception can be induced with a sequence of static images. These effects are known as apparent motion (Exner, 1875; Wertheimer, 1912). When presented with video, we perceive a continuous motion, and not a succession of 24 image frames. There are limitations to the velocity that can be perceived,

as is evidenced by the wagon wheel effect. Apparent motion can also be induced, simply by gradually changing the intensity in a pattern. The system then computes a temporal change in the intensity and perceives motion, the so-called reverse phi effect (Anstis, 1970; Anstis and Rogers, 1986). Furthermore, biological systems do not, as suggested by video technology, compute image motion directly from the difference between the frames in a video sequence. Instead the neurons in the early stages of processing produce responses whenever the change of intensity at an image point exceeds some value (Laughlin, 1981; Srinivasan et al., 1982). This difference in response time can account for the difference in velocity perceived in motion sequences on patterns with significantly different contrast in different regions, such as the stepping feet illusion (Anstis, 2003).

Visual motion analysis encompasses two distinct processes: the computation of image motion, and the interpretation of the image motion in terms of the 3d scene and 3d movement. The early interpretation involves computing the observer's own motion, or ego-motion, the segmentation of the scene into surfaces at different depth, detection of objects that move in the scene and the estimation of their movement. The observer's own motion is due to a combination of the body, head, and eye movements. For some of these, other measurements from the inertial sensors are available and combined with the image measurements. Computational approaches treat the whole process strictly in two successive steps, first computing image motion then interpreting it. The many feedback loops found from neural studies indicate that biological systems do not implement it in such a strict bottom-up fashion, but rather the processes are intertwined and solved iteratively. How these feedback loops work is still not clear. But this does not matter for our model here. It is important only to understand that image motion contributes to ego-motion and segmentation, but the three processes are not independent but shape each other.

3 Image Motion Estimation:

Image motion is obtained from the changing image patterns on the retina over time. As a representation for image motion we use the concept of the *optical flow.* The optical flow vector associated with an image point represents the point's instantaneous movement. The corresponding field of all measurements (a vector at every image point), the so-called optical flow field, represents an approximation of the projection of the 3d scene points' movement on the image. Computational considerations as well as biological measurements suggest that optical flow is computed in two stages (Adelson and Movshon, 1982).

In a first stage, considering only local image measurements in a small spatio-temporal region, the velocity component perpendicular to linear features is computed. The situation is illustrated in Fig. 1a. The velocity vector of a line segment viewed through a small aperture is inherently ambiguous, as it is consistent with any vector falling on the constraint line (Wallach, 1935). Only the velocity component perpendicular to the line feature in the direction of the motion is well defined. This component of the image motion is referred to as *normal flow*, and the ambiguity is referred to as the *aperture problem* (Marr and Ullman, 1981). In order then to derive in a second stage the complete two-dimensional optical flow, one-dimensional normal flow measurements from line features in different directions in a neighborhood need to be combined, and the optic flow estimated from them.

When we compute optical flow, we treat the image intensity as a smoothly varying function of space and time, and we derive the optical flow from the changes in the image intensity. Either we use filters that approximate the derivatives in image domain or we use filters tuned to certain frequencies

of space and time, such as biological plausible motion filters. It is generally considered that image motion can be estimated quite well at locations in the image where the assumptions hold; these are textured regions on smooth surfaces. It is also known, that there are problems at the locations where the scene changes, i.e. at the boundaries of 3d surfaces, because there the image intensity function is discontinuous. At these locations occlusions make it difficult to obtain accurate normal flow, and the estimation of optical flow is even more challenging (Horn and Schunck, 1981).

In this chapter we show, that even within areas of smooth flow, that is the regions corresponding to smooth textured 3d surfaces, the computation of flow poses a problem, and both normal flow and optical flow are affected, as summarized next.

The estimation of normal flow is erroneous for certain signals, because real-time systems estimate temporal changes using causal filters. That is, motion is found from the response of spatio-temporal filters, with the filters tuned to spatial frequencies being symmetric, but the filters tuned to temporal frequencies being asymmetric. This is because real-time systems receive as input data from the present and the past. The so-called causal filters used for processing this data are asymmetric with greater weight given to recent input than older input, because otherwise the processing would be much delayed. Such causal filters mis-estimate the image motion in asymmetric image signals for certain spatial frequencies.

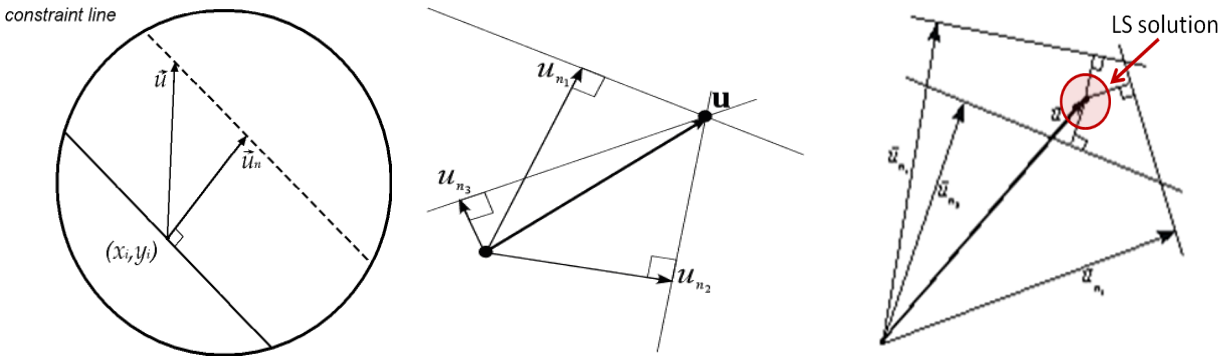
The estimation of optical flow cannot be accurate because of statistical reasons. Like any real system, our vision has to deal with noisy data. The local normal flow measurements can be estimated only within some range of the actual value, in other words they are noisy. Because of this noise, the estimation of optical flow using as input normal flow is statistically biased. In other words, the expected value and the actual value are different. From a statistical viewpoint, to correct for the bias, would require knowledge of the noise parameters. The noise parameters would have to be estimated from the data, but they are difficult to obtain, because the parameters are not static, but instead they change in unpredictable and complex ways, as the scene, the viewing conditions, and the lightening changes. There always is bias in the estimation, but in most situations the bias is insignificant. It depends on the texture in the image, and can be large for regions with a dominant gradient distribution.

The next two sections discuss the mis-estimation of image motion in more detail and how it can account for optical motion illusions. Section 4 discusses the challenges in the estimation of optical flow, and Section 5 discusses the challenges in normal flow estimation. Section 6 concludes with a discussion on the role of early motion processes and their interaction with other processes of static cues.

4 Bias in optical flow estimation

There are two dominant models for the computation of optical flow in the literature: gradient-based methods and frequency based methods. Both methods derive optical flow in a two-stage process: first information about the one-dimensional motion component of local edges or single spatial frequencies is obtained; then, the individual measurements within some neighborhood are combined into an estimate of optical flow. The two methods are faced with similar noise issues and can be given a similar mathematical modeling. [1]

[1] A third class of methods used in the computational literature, models optical flow as correlation of regions followed by interpolation, and does not proceed in the same two steps and requires a somewhat different mod-



(a) (b) (c)

Fig. 1. (a) The aperture problem: Observing a line feature (solid line) through a small aperture, we cannot determine where exactly the point (x_i, y_i) moves to. We can only obtain a constraint line (dashed line), which provides the component of flow perpendicular to the line. (b) If there is no noise, the constraint lines of different image points with their different gradient directions will intersect in one point. (c) In the presence of noise, we estimate the optical flow by minimizing some distance; in the case of LS estimation this is the average distance to the constraint lines.

The monochromatic light distribution on the retina can be described as a function $I(x,y,t)$ that specifies the intensity at a point (x,y) at time t. Let us denote the instantaneous motion of a point in the image by a translation of velocity $\vec{v}$ with horizontal and vertical speed components (u,v). Assuming that the intensity at a point over a small time interval does not change, that is

$$I(x,y,t) = I(x-ut, y-vt, 0),$$

we obtain from a first order Taylor expansion of this equation the following constraint relating the spatial and temporal derivatives I_x, I_y and I_t and the image motion

$$I_x u + I_y v = -I_t. \tag{1}$$

This equation, called the *optical flow constraint equation* (Horn and Schunck, 1981), defines at a point the component of flow in the direction of the spatial gradient (I_x, I_y), or in geometric terms, it defines the component of the flow perpendicular to the edge (as illustrated in Fig. 1a), and this component is termed the normal flow. Similarly, taking the three-dimensional Fourier transform of equation 1, we obtain (Watson and Ahumada, 1985; Adelson and Bergen, 1985)

$$u\omega_x + v\omega_y = -\omega_t, \tag{2}$$

where ω_x, ω_y denote the spatial frequencies and ω_t the temporal frequency. We can also interpret this equation as, given a dominant spatial frequency at an image point, we know the flow component parallel to the direction of this frequency. The frequencies can be computed from data of a spatio-temporal volume using filters tuned to certain frequencies.

To derive the optical flow, a second mathematical constraint has to be formulated. Most computational methods assume that the flow varies smoothly. The very simplest assumption is that the flow is constant within a neighborhood. This is called the *intersection of constraints (IOC)* model. We will use this model to keep the exposition simple.

eling (Fermüller et al., 2001).

Let us consider that in a neighborhood we have n measurements. If considering a gradient based approach, each measurement i provides an equation of the form (eq. 1), and we obtain an over-determined system of equations, which we write as:

$$A\vec{u} = \vec{b}, \tag{3}$$

where A denotes the $n \times 2$ matrix of spatial gradients (I_{x_i}, I_{y_i}) and $\vec{b}$ the n-dimensional vector of temporal derivatives (I_{t_i}). Similarly, if considering a frequency based approach (eq. 2), A and $\vec{b}$ are the spatial and temporal frequencies $(\omega_{x_i}, \omega_{x_i})$ and ω_{t_i}. Each equation in (3) provides a constraint line, and in principle, from two measurements we can solve for the optical flow. If all the measurements (I_{x_i}, I_{y_i}) and I_{t_i} were perfect, all constraint lines would intersect in a point. However, because they are noisy, we have to use an estimator to find a solution that is best with respect to some criterion. Standard least squares (LS) estimation finds the point that has minimum normal distance to all the constraint lines (see Fig. 1b and c). The LS solution of (eq. 3) for the flow yields the estimate

$$\vec{u} = (A^T A)^{-1} A^T \vec{b}, \tag{4}$$

where A^T and A^{-1} denote the transpose and the inverse of matrix A respectively.

The observed data is always noisy, or we say the observations $A = (a_{1_i}, a_{2_i})$ and (b_i) are corrupted by errors. We use unprimed letters, primed letters and δ's to denote the estimates, actual values and errors, respectively. We can say that the observations are composed of the true values (A', b') and the errors $(\delta A, \delta b)$, and rewrite equation 3 as

$$(A + \delta A)\vec{u} = (\vec{b} + \delta\vec{b}). \tag{5}$$

It is well known, that least squares estimation generally is biased (Fuller, 1997; van Huffel and Vandewalle, 1991). It provides an unbiased solution if noise occurs in variable $\vec{b}$ only, that is in our case the length of the normal flow, but it is biased if there is noise also in the explanatory variables A, in our case, the direction for the normal flow. Let us clarify, the concept of statistical bias means, that if we had many, many noisy measurements to estimate the unknown parameter, then the expected value of our estimate (in the limit) would be different from the true value.

Under some simplifying assumptions (identical and independent random variable δA and $\delta\vec{b}$ with zero mean and variance σ^2) , the estimate $E(\vec{u})$ and the true value $\vec{u}'$ are related as (Fuller, 1997)

$$E(\vec{u}) = \vec{u}' - \sigma^2 (\lim_{n\to\infty} (\frac{1}{n} A'^T A'))^{-1} \vec{u}', \tag{6}$$

which implies that the estimate is asymptotically biased. The bias amounts to $\sigma^2(\lim_{n\to\infty}(\frac{1}{n}A'^T A'))^{-1}\vec{u}'$. It depends on the amount of noise, σ^2, the texture, which is described by matrix $M' = (A'^T A')$, and the relationship between the actual flow and the texture. An analysis of the bias term reveals, that large variance in δA, an ill-conditioned A', or a $\vec{u}'$ which is oriented close to the eigenvector of the smallest eigenvalue of $A'^T A'$ all could increase the bias and push the LS solution away from the real solution. Generally it leads to an underestimation of the parameters.

To obtain intuition why there is underestimation, consider the simpler problem of estimating the one-dimensional unknown x from a set of linear equations of the form $ax = b$. As an example, consider two measurements with a symmetric noise distribution. Let these measurements be $a_1 = a' + \delta a$,

$a_2 = a' - \delta a$. (Noise symmetric in b doesn't have an effect on the bias, so we do not need to consider it in our example; thus $b_1 = b_2 = b'$.) The LS solution of the linear equation amounts to $x = \frac{ab}{a^2}$, and thus

$$x = \frac{\sum_{i=1...2} a_i b_i}{2\sum_{i=1..2} a_i^2}. \tag{7}$$

Substituting for the values of a_1, a_2, b_1, b_2, we obtain

$$x = \frac{(a' + \delta a + a' - \delta a)b'}{(a' + \delta a)^2 + (a' - \delta a)^2} = \frac{a'b'}{(a'^2 + 2(\delta a)^2)} > \frac{a'^2}{a'b'}. \tag{8}$$

This intuitively explains the bias. Basically, the bias originates from the quadratic term a^{-2}, and it is due to the variance of δa. In the two-dimensional case a^{-2} becomes $(A^T A)^{-1} = ((A' + \delta A)^T)(A' + \delta A))^{-1}$ and can be obtained from a second order Taylor expansion of equation 4.

The information about the bias is encoded in $M' = A^T A$, the matrix of spatial gradients, which describes the texture in a patch. In the case of a uniform distribution of the image gradients the bias is solely in the length of the computed optical flow; there is an underestimation. Consider two dominant gradient directions. The signal to noise ratio is smaller in the direction of fewer measurements. Thus there is more underestimation in the direction of fewer measurements and less underestimation in the direction of more measurements. As a result the estimated flow is biased downward in size and biased toward the major direction of the gradients in the patch.

4.1 A note on estimation

We have shown that least square estimation is biased. One may ask, is the vision system really doing least squares? Couldn't it do better? Looking at the statistical problem, least square estimation is only one of many possible estimators. The question then is, are there other estimators that do not suffer from this problem? An elaborate discussion on the statistics of optical flow estimation can be found in (Fermüller et al., 2001). The analysis includes all categories of image motion methods: gradient based, energy based, and correlation methods, it includes different, more elaborate models for combining one-dimensional measurements into two-dimensional velocity estimates in a neighborhood including linear and nonlinear methods, and it includes many statistical estimators. The analysis shows that all methods suffer from bias, and for many statistical estimators the bias is an underestimation. The reason why it is so difficult to correct for the bias, is that this would require knowledge of the noise parameters. These parameters would need to be estimated from the data, but usually there is not enough data available about the value of the noise. It might be possible if the parameters were static, but instead they change in unpredictable ways with the change of environment, the lighting and viewing conditions.

Statistical estimators have to deal with two components: one is bias, the other is variance; and there is a trade-off between the two. Generally an estimator correcting for bias increases the variance while decreasing the bias. Thus, if in some cases there is data to extract the statistics of the noise, for example when we look at the same static scene for some time, theoretically the best thing to do is to partially correct the bias. This does not change the form of the bias, but decreases its amount. We believe that the human vision system, when it has access to enough data, follows this theoretically best approach. Two observations make us believe so. First, illusory perception in many optical illusions weakens after extended viewing, in particular when subjects are asked to fixate (Helmholtz, 1962; Yo

and Wilson, 1992). In these cases the noise parameters stay fixed, the vision system can reasonably estimate them well and partially correct. Second, in experiments (Ji and Fermüller, 2006) we varied the variance by changing the density of the texture in a pattern. Theoretically, a better correction and thus better estimation is possible for a texture that is dense and thus has small variance. The experiments found that our perception is consistent with this hypothesis.

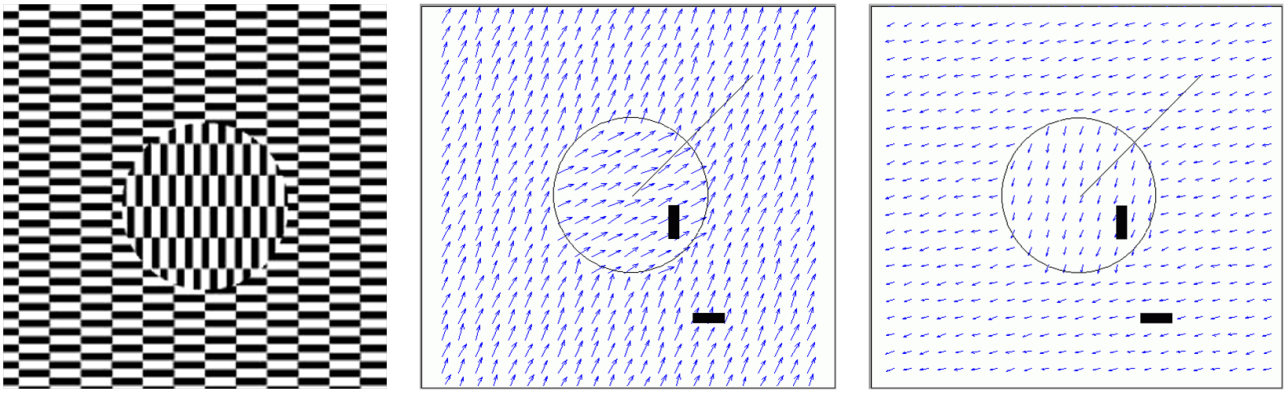
Finally, we think that the bias not only is a problem of flow estimation, but many other visual computations which amount to estimations processes. Early visual processes estimate features, such as lines, points and image motion (Fermüller and Malm, 2004). All these processes are biased, and as a result the locations of features are perceived erroneously. These effects can account for many classically optical geometric illusions, as well as the motion illusion explained next. The processes of computing the shape of the scene from image features, called *shape from X* computations are also estimation processes. The bias can account for the erroneous estimation of shape (an underestimation of slant) found in many psychophysical experiments (Ji and Fermüller, 2006) as well as illusions of shape, that we created using the insight on the bias.

4.2 Ouchi illusion

The bias provides an explanation for many motion illusions, including the well-known Ouchi pattern. This striking illusory pattern created by the graphic artist H. Ouchi, and introduced as illusion to the research community by Spillmann et al. (1986) consists of two black and white rectangular checkerboard patterns oriented in orthogonal directions – a background orientation surrounding an inner ring (Fig. 2a). Small retinal motions, or slight movements of the paper, evince a segmentation of the inset pattern and motion of the inset relative to the surround. The illusion occurs for a variety of viewing distances and angles. Some observers report an apparent depth discontinuity, with the center floating as it moves above the background (Spillmann et al., 1993). In a nutshell, the cause of the illusion is differently biased flow vectors in the two patterns (Fermüller et al., 2000), as follows.

The tiles used to make up the pattern are longer than they are wide. Thus, in any small region there are many more normal flow measurements in one direction than the other. Since the tiles in the two regions of the figure have different orientations, the estimated regional optical flow vectors are different (they have different bias). The global motion of the pattern is computed from the data in the surround. When looking at the static pattern under free viewing conditions, small retinal eye movements cause the motion, which are estimated from the periphery. The difference between the bias in the inset and the bias in the surrounding region is interpreted as motion of the ring. An illustration of the estimated flow and the difference of the estimated and veridical flow is given in Fig. 2b and c for the case of motion along the first meridian (to the right and up). In addition to computing flow, the visual system also performs segmentation, which is why a clear relative motion of the inset is seen.

The theory can account for variations in the pattern. In figure 2 the ratio of the two sides of the rectangle is 4 : 1; a smaller ratio is predicted to lead to smaller bias as has been perceived. The bias also depends on the direction of movement, it is predicted to be largest for a diagonal movement and smallest for movements parallel to the sides, as can also be perceived. Khang and Essock (1997) created variations of the Ouchi pattern by replacing the periodic rectangular with other periodic functions, that are "smoother" and have more local gradient orientations. The bias model predicts that with the spreading of directions, the amount of bias in the estimated flow decreases, which explains the decrease in the perceived illusory motion found in the experiments.



(a) (b) (c)

Fig. 2. (a) A variant of the Ouch illusion. (b) Estimated optical flow field (c) The error vector field the difference between the estimated and the veridical motion. The line from the center is the direction of the veridical motion.

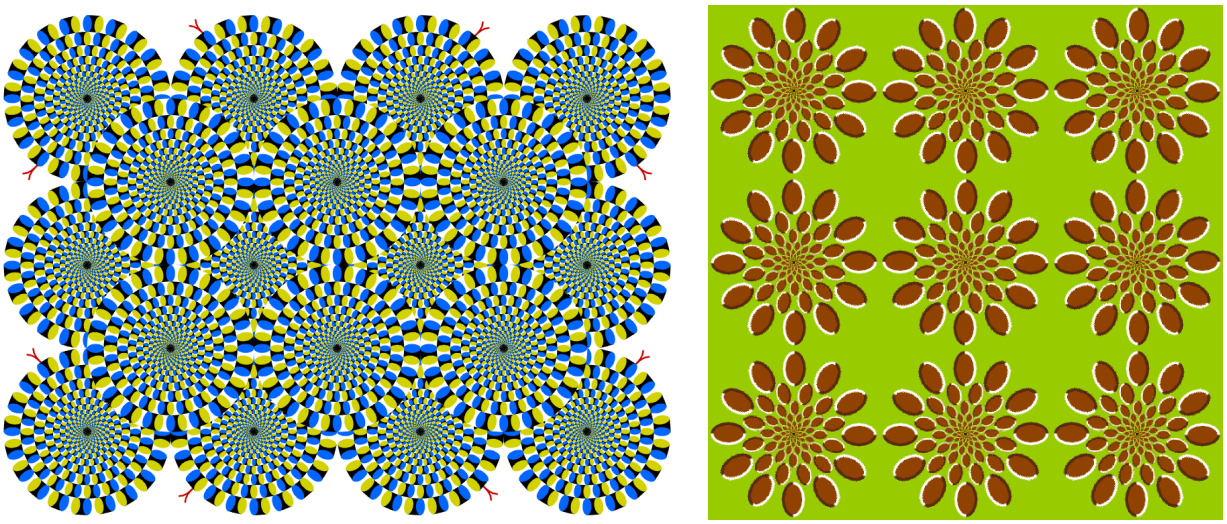
Many other illusory percepts can be explained by the bias, such as the erroneous perception of motion in plaid patterns, the Wheel illusion (Pinna and Brelstaff, 2000), the perceived motion of contracting and expansion in rotating spirals, and the rigid vs no-rigid motion perceived in translating sinusoids of high and low amplitude respectively (Fermüller, 2005).

5 Errors in normal flow estimation due to causal filtering

Next we look at the estimation of normal flow in more detail, and we show that for certain signals and certain image resolutions, one-dimensional image motion is mis-estimated. These signals are asymmetric at some resolution. Take a look at the striking illusion "Rotating Snakes" (Fig. 3a) (Kitaoka, 2003). Most observers experience very strong illusory movement when viewing this and similar patterns, such as Fig. 3b. The illusory movement is experienced under free viewing conditions when one moves the eyes and is perceived in non-central vision. For example, while one looks at the upper left of the pattern, motion is perceived in its central and lower right part. From this we see, an essential aspect of the effect is that it depends on the image resolution.

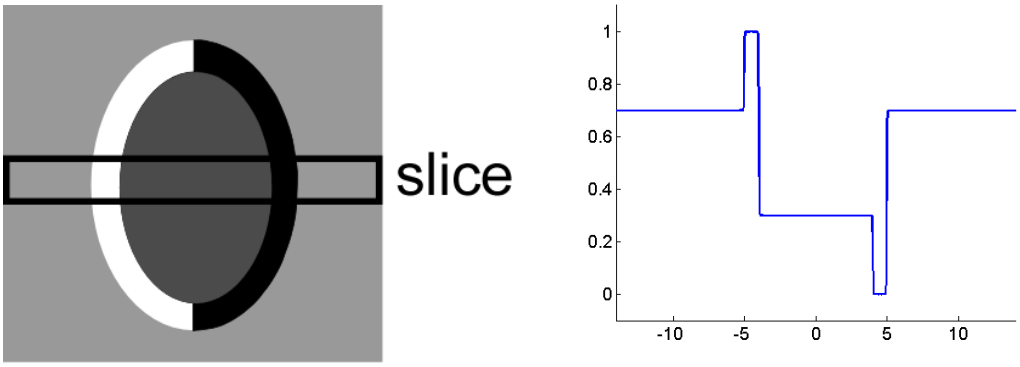
The patterns are composed of image patches which have an asymmetric intensity profile. For example, consider a narrow slice in the middle region of one of the ovals in Donguri, as shown in Fig. 4a. (The Japanese word donguri translates to acorn.) Its monochromatic intensity image can be described as a white and a dark bar (the boundaries of the oval) next to different shades of gray. Referring to Fig. 4b, 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 on the right than on the left. Thus, at the two bars the changes of intensity in the right and left neighborhood are different. Informally, we say that the pattern is asymmetric (at the scale of the size of the bars). 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) (Ashida and Kitaoka, 2003).

The analysis is performed using filters as it allows us to analyze the signal under a continuous change in scale. However, a gradient based approach, observing the changes under smoothing with gradually increasing smoothing and difference kernels will lead similar results.



(a) (b)

Fig. 3. (a) Rotating Snake. (b)Variation of Donguri pattern. In peripheral vision most observer experience rotary movement in both patterns. The direction in the circular arrangements alternates, with counter-clockwise direction in the upper left.



(a) (b)

Fig. 4. (a) Slice through a patch in the Donguri pattern. (b) Its intensity profile.

To simplify the analysis, we consider a one-dimensional image with bar-like structure parallel to the vertical dimension and the motion component perpendicular to the bars. Thus, we have a two-dimensional function $I(x,t)$, and the image motion constraint (eq. 2) becomes

$$u\omega_x = -\omega_t, \tag{9}$$

defining a line in the two-dimensional frequency space. The velocity u is obtained from the ratio of the temporal and spatial frequency, and amounts to

$$u = -\frac{\omega_t}{\omega_x}. \tag{10}$$

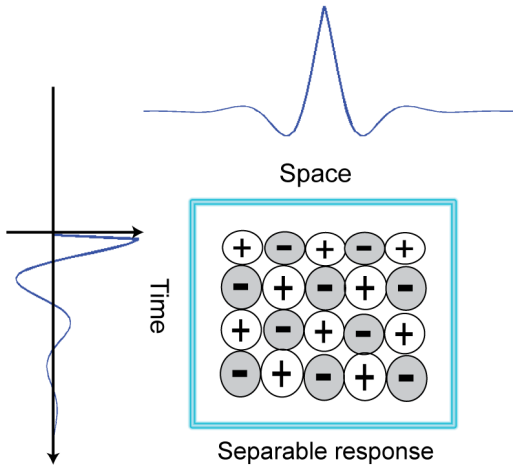


Fig. 5. Illustration of biological implementation of spatio-temporal filter (similar to Fig.6 of (Adelson and Bergen, 1985)). The spatial and temporal impulse responses are shown along the margins. Their product is shown schematically in the center.

5.1 The filter

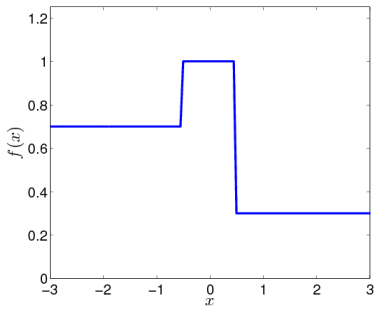
The spatio-temporal energy filters for extracting motion are separable in space and time. This just means that the filters can be created as the product between a spatial and a temporal filter. For the analysis this means that the spatio-temporal signal may first be convolved with the spatial filter and the result may then be convolved with the temporal filter. We follow the common formulation of modeling a filter for detecting the local frequency ω_0, as a complex function

$$g(y) = p(y) \cdot exp(2\pi i \omega_0 y). \tag{11}$$

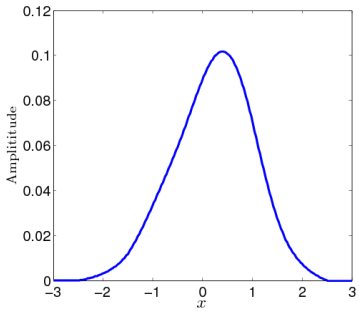
$exp(2\pi i \omega_0 y)$, called the carrier function, is a complex sinusoid for detecting the signal's component of frequency ω_0, and $p(x)$, called the envelope function, localizes the sinusoid in image space. (Complex functions are needed so we can extract motion independent of the phase of the signal, that is independent of the position of the signal within the receptive field at certain time, and independent of the sign of the contrast. (Adelson and Bergen, 1985; Watson and Ahumada, 1985)). Fig. 5 illustrates the impulse response function of a biological plausible filter with symmetric envelope in the spatial domain and asymmetric envelope in the time domain (Adelson and Bergen, 1985; Burr and Morrone, 1993). In our model the envelope is a Gaussian in the spatial domain leading to a Gabor filter, and a Gamma probability density function in the temporal domain (Chen et al., 2001; Shi et al., 2004).

5.2 The effect of filtering on Donguri

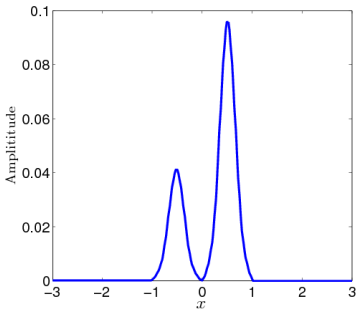
Figure 6 illustrates the effect of spatial filtering on Donguri. At frequencies ω_x larger than the width of the bar, the Gabor filter detects the two edges at the left and right of the bar (Fig. 6b). At frequencies significantly smaller than the width of the bar, the Gabor detects the bar (Fig. 6c). 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. 6d). We call these frequencies the "critical frequencies".



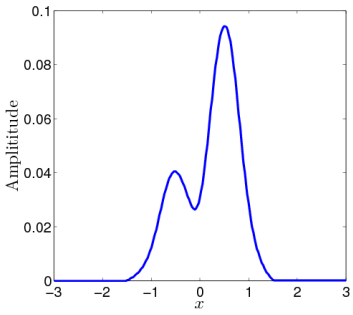
(a)



(b) $\omega_x = \frac{1}{2}$



(c) $\omega_x = 2$



(d)$\omega_x = 1$(”*critical frequencies*”)

Fig. 6. (a) Bar in Donguri pattern. (b-d) Amplitude of bar filtered with Gabor of different frequencies ω_x.

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. The filtered signal should have a local dominant frequency of ω_x, but, because of the “hat” profile of its Gaussian envelope, this will not always be the case. For the critical frequencies the value is quite different from ω_x.

When now estimating on the asymmetric filtered signal, image motion with asymmetric temporal filters, left and right motion are estimated of different value. We can intuitively understand this from the shape of the signals: We convolve the asymmetric spatially filtered signal (Fig. 6d) with larger weight on the right and smaller weight on the left with a temporal filter that in the one case has larger weight on the left and smaller weight on the right, and in the other case with a filter with larger weight on the right and smaller weight on the left (the mirror-reflection). As a result, for the critical frequencies, motion to the left $(u = -1)$ leads to larger velocity estimates than motion to the right $(u = 1)$. In some more detail, Fig. 7 shows the local estimated velocity (as full, green line) at every point on the bar. The corresponding amplitude is shown as dot-dashed, red line, and the amplitude of the spatially filtered signal is shown as dashed, blue line (in the spatial domain). Because of interaction of the regions under the two peaks with each other during temporal filtering, the local velocity (at a single point) varies significantly along the signal. Most significant, there is overestimation of velocity at the right peak for left motion, and underestimation of temporal energy at the left peak for right motion. As a result the average velocity (average over the signal) is over-estimated for left motion, and underestimated for right motion.

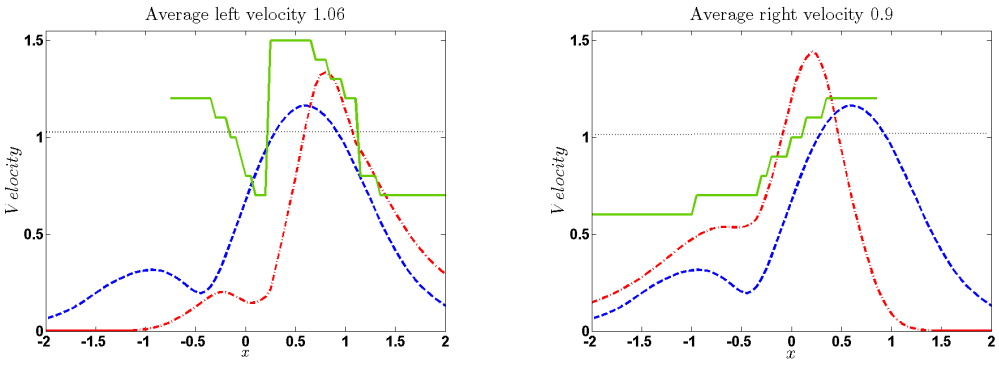


Fig. 7. Velocity estimation at critical frequencies obtained from a simulation. The dashed blue line denotes the amplitude of the spatially filtered signal. The full green line denotes the local estimated velocity and the dot-dashed red line denotes the corresponding scaled amplitude. Both amplitudes have been scaled to allow for better visualization. The estimated average velocity is larger for left than for right motion.

5.3 Explaining the illusion

The complete explanation of the illusions is as follows. The motion is caused by involuntary fixational eye movements. Work by Murakami et al. (2006) implicates drift eye movements [2]. 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 the parameters of the rigid motion and compensates for the image motion field, i.e. the images are stabilized (Murakami, 2004; Murakami and Cavanagh, 1998). Even for asymmetric signals, the vision system can estimate a quite accurately the 3D rigid motion using the average of the motion vectors in the patterns. However, as shown above, a mis-estimation occurs at certain locations in the image. Local one-dimensional motion signals are estimated erroneously, and as a result the two-dimensional image motion estimated as shown in section 4 (using eq. 2), is erroneous as well. 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 are integrated over time and space causing the perception of illusory motion in the image.

To evaluate quantitatively the predictive power of the model we performed an experiment with human subjects (Fermüller et al., 2010). Three different signals of bar-like structures, giving rise to differently strong illusory percepts, were compared in a nulling experiment with opposing real motion. The different signals were arranged on concentric circles shown to naive subjects, who were instructed to adjust the speed of motion in a video till they experienced a stationary pattern. The model was shown to predict well the relative motion speed in the different signals. The model also was shown to account for a series other illusions, experienced in central and peripheral vision, which Kitaoka and Ashida (2004) call central drift illusions.

Finally, we summarize existing explanations of this illusion, which we broadly classify into two types. The first type, which includes our model and work by (Ashida and Kitaoka, 2003) suggests that

[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.

small eye-movements cause retinal motion, and the illusion arises due to asymmetric temporal processing. Our model mathematically explains the reason for the erroneous estimation, and introduces the concept of the dependence of the estimation at different resolutions of the pattern. The second kind of theory, originating from Faubert and Herbert (1999), suggests that temporal differences in luminance processing produce a motion signal. The theories differ in how this signal is produced. The main idea is that high-contrast regions of the image are transmitted faster than contrast low-contrast regions leading to a motion from high-contrast to low-contrast areas (Conway et al., 2005). Backus and Oruç (2005) suggest that differently strong contrasts and intensities cause different neural response curves over time. Their model introduces the effect of adaptation which can account for the smooth perception under fixation over a few seconds. This effect may likely exist in addition to the one discussed here.

The illusory perception in the snake pattern is perceived under free viewing conditions and also when flashing the pattern, and many theories seek to explain these effects with a single cause. It is easy to understand that flashing will produce a reverse phi motion Conway et al. (2005). – Reverse phi motion is an image motion effect caused by reversing the contrast in some frames of a video sequence. However, in our opinion the illusory effect under free viewing and the effect when flickering the pattern are not the same. In support of this, we observed that for the reduced experimental stimuli the latter is experienced much stronger than the former, and it is experienced even by observers who do not experience the effect under free viewing.

6 Summary and Discussion

This chapter discussed that for reasons of statistics and signal processing, it is not possible to estimate correct image motion for all signals. First, temporal image motion filters are causal, i.e. they use data from the past, but do not use data from the future. The asymmetric asymmetry in these filters leads to erroneous estimation of local one-dimensional image motion measurements for asymmetric signals at certain scale. Second, the estimation of two-dimensional image motion using as input one-dimensional motion signals is statistically biased. Because there is noise in the spatio-temporal signal, the estimation cannot be without error. We argue that these limitations in the estimation of flow are integral also to human vision and are the main cause of many optical illusions. We tested our models quantitatively, and used them to create new illusory percepts.

Estimation of image motion estimation is the first step in visual motion analysis. The local image motion signals are input to further visual processes. The very basic motion processes include the estimation of our own motion, that is our body motion and eye movements, the movement of objects in the scene, and the segmentation of the scene into different surfaces. While causal filters and bias in estimation create local erroneous motion signals in the motion illusions discussed above, the strong perception of parts of the image moving with a rotating, sliding, or expanding movement is due to these further processes. But how the interaction between the different processes is implemented, is still a topic of research. In our opinion there are two major questions where research on optical illusions may help us gain insights.

First, what is the role of motion analysis in the estimation of self motion and eye-movements? When we move, the image motion on our eyes are due to our body movements, our voluntary and involuntary eye-movement, and the movements of objects. The inertial sensors also provide information. But how are the different sources integrated, and the different components of movement sep-

arated. Under normal circumstances, the images on our eyes are stabilized; we have a stable model of the scene and do not experience micro-movements. As an example of studying an optical illusion, consider the snake illusion. While it can be shown that there are errors on any image motion in asymmetric signals, the illusory perception in the snake illusion has been shown to be due only to the motion signal from drift movements (Murakami et al., 2006; Kuriki et al., 2008). We speculate that drift movements allow for a better temporal integration of the motion signal when compared to signals from other movements. The drift motion is computed from the local motion signals over the whole visual field. Then the drift is discarded, and the image signals over a time interval are integrated. This is computationally feasible, because the drift motion is mostly a rotation and does not depend on the structure of the scene. By fitting to the whole image motion field a rotational motion field, which only depends on three parameters, local motion vectors can be estimated very accurately and reliably. On the other hand, head motions and scene motions also involve translation, and the image motion field then depends on the scene. Therefore, local motion estimation cannot be that accurate, and integration over a time interval is more difficult.

Second, how do the different processes of early motion perception interact? Insight from computational and biological studies indicate that biological systems implement the estimation of image motion, self-motion and motion segmentation not in a bottom-up fashion, but rather solve them iteratively. As the system moves, it keeps updating its world model. We would argue, that first only approximations of image motion and rough segmentation are derived, and self motion is computed. Then estimates of 3d motion and segmentation help with estimating better image motion and in turn more accurate segmentation at the next time instance. Furthermore, at the level of segmentation, image motion is combined with static cues such as contour, color and texture to obtain the boundaries between surfaces. As an example of using an illusion to gain insight, consider the Leviant illusion (Zeki et al., 1993). We hypothesize that the main reason for this illusion lies in the interaction between the three processes of image motion, ego-motion and segmentation, and in (Fermüller et al., 1997) we showed that we can create variants of this illusion from patterns which have a special role in ego-motion estimation.

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.

Adelson, E. H., Movshon, J. A., December 1982. Phenomenal coherence of moving visual patterns. Nature 300, 523–525.

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

Anstis, S., 2003. Moving objects appear to slow down at low contrasts. Neural Networks 16, 933–938.

Anstis, S., Rogers, B., 1986. Illusory continuous motion from oscillating positive-negative patterns: implications for motion perception. Perception 15, 627–640.

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

Backus, B. T., Oruç, I., 2005. 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.

Chubb, C., Sperling, G., 1988. Drift balanced random dot stimuli; a general basis for studying non fourier motion. Journal of the Optical Society of America A 5, 1986–2007.

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.

Exner, S., 1875. Experimentelle untersuchung der einfachsten psychischen processe. Pflugers Arch. Physiol. 1, 403–472.

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

Fermüller, C., 2005. http://www.cfar.umd.edu/∼fer/optical/.

Fermüller, C., Ji, H., Kitaoka, A., 2010. Illusory motion due to causal time filtering. Vision Research 50 (3), 315–329.

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

Fermüller, C., Pless, R., Aloimonos, Y., 1997. Families of stationary patterns producing illusory movement: Insights into the visual system. Proc. Royal Society, London B 264.

Fermüller, C., Pless, R., Aloimonos, Y., 2000. The Ouchi illusion as an artifact of biased flow estimation. Vision Research 40, 77–96.

Fermüller, C., Shulman, D., Aloimonos, Y., 2001. The statistics of optical flow. Computer Vision and Image Understanding 82 (1), 1–32.

Fuller, W., 1997. Estimated true values for errors-in-variables models. In: van Huffel, S. (Ed.), Recent Advances in Total Least Squares Techniques and Errors-in-Variables Modeling. SIAM.

Gillam, B., 1998. Perception and Cognition at Century's End. Academic Press, Ch. Illusions at Century's End, pp. 95–136.

Grossberg, S., Todorovic, D., 1988. Neural dynamics of 1-d and 2-d brightness perception: A unified model of classical and recent phenomena. Perception and Psychophysics 43, 241–277.

Helmholtz, H. L. F. V., 1962. Treatise on Physiological Optics. Vol. III. Dover, New York, translated from the Third German Edition by J. P. C. Southall.

Horn, B. K. P., Schunck, B., 1981. Determining optical flow. Artificial Intelligence 17, 185–203.

Ji, H., Fermüller, C., 2006. Noise causes slant underestimation in stereo and motion. Vision Research 46 (19), 3105–3120.

Khang, B.-G., Essock, E. A., 1997. Apparent relative motion from a checkerboard surround. Perception 26, 831–846.

Kitaoka, A., 2003. http://www.ritsumei.ac.jp/∼ akitaoka/index-e.html.

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

Kuriki, I., Ashida, H., Murakami, I., Kitaoka, A., 2008. Functional brain imaging of the rotating snakes illusion by fmri. Journal of Vision 8 (10), 1–10.

Laughlin, S., 1981. Neural principles in the peripheral visual systems of invertebrates. In: Autrum, H. (Ed.), Handbook of Sensory Physiology. Vol. 7. Springer, pp. 133–280.

Marr, D., 1982. Vision. W.H. Freeman, San Francisco, CA.

Marr, D., Ullman, S., 1981. Directional selectivity and its use in early visual processing. Proc. Royal

Society, London B 211, 151–180.
Mingolla, E., 2003. Neural models of motion integration and segmentation. Neural Networks 16 (5/6), 939–945.
Morrone, M., Burr, D., 1988. Feature detection in human vision: a phase dependent energy model. Proc. Royal Society, London B , 221–245.
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.
Neumann, H., Mingolla, E., 2003. Contour and surface perception. In: Arbib, M. (Ed.), Handbook of brain theory and neural networks. Vol. II. MIT Press, Cambridge, MA, pp. 217–276.
Palmer, S. E., 1999. Vision Science: Photons to Phenomenology. MIT Press, Cambridge, MA.
Pinna, B., Brelstaff, G. J., 2000. A new visual illusion of relative motion. Vision Research 40 (16), 2091–2096.
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.
Spillmann, L., Heitger, F., Schuller, S., 1986. Apparent displacement and phase unlocking in checkerboard patterns. In: 9th European Conference on Visual Perception. Bad Nauheim.
Spillmann, L., Tulunay-Keesey, U., Olson, J., 1993. Apparent floating motion in normal and stabilized vision. Investigative Ophthalmology and Visual Science, Supplement 34, 1031.
Srinivasan, M., Laughlin, S., Dubs, A., 1982. Predictive coding: a fresh view of inhibition in the retina. Proc. Royal Society, London B 216, 427–459.
van Huffel, S., Vandewalle, J., 1991. The Total Least Squares Problem: Computational Aspects and Analysis. SIAM.
Wallach, H., 1935. Über visuell wahrgenommene Bewegungsrichtung. Psychologische Forschung 20, 325–380.
Watson, A. B., Ahumada, A. J., 1985. Model of human visual motion sensing. Journal of the Optical Society of America 2, 322–342.
Wertheimer, M., 1912. Experimentelle Studien über das Sehen von Bewegung. Z. Psychol. 61, 161–265, translation in Shipley, T. (Ed.), *Classics in Psychology*, Philosophical Library, New York, 1961.
Yo, C., Wilson, H. R., 1992. Moving 2D patterns capture the perceived direction of both lower and higher spatial frequencies. Vision Research 32, 1263–1270.
Zeki, S. M., Watson, J. D. G., Frackowiak, R. S. J., 1993. Going beyond the information given: The relation of illusory visual motion to brain activity. Proc. Royal Society, London B 252, 215–222.