Phase-differencing in Stereo Vision
Solving the Localisation Problem

J. M. H. du Buf, K. Terzic and J. M. F. Rodrigues
Vision Laboratory, LARSyS, University of the Algarve, 8005-139 Faro, Portugal

Keywords: Complex Gabor Filters, Phase, Disparity, Lines, Edges, Localisation.

Abstract: Complex Gabor filters with phases in quadrature are often used to model even- and odd-symmetric simple cells in the primary visual cortex. In stereo vision, the phase difference between the responses of the left and right views can be used to construct a disparity or depth map. Various constraints can be applied in order to construct smooth maps, but this leads to very imprecise depth transitions. In this theoretical paper we show, by using lines and edges as image primitives, the origin of the localisation problem. We also argue that disparity should be attributed to lines and edges, rather than trying to construct a 3D surface map in cortical area V1. We derive allowable translation ranges which yield correct disparity estimates, both for left-view centered vision and for cyclopean vision.

1 INTRODUCTION

Unquestionably, the phase of Gabor filters provides useful information for stereo disparity. It is also biologically plausible, because simple cells in the primary visual cortex (area V1) are often modelled by complex Gabor filters with phases in quadrature. The cortical structure in hypercolumns brings information of the left and right eyes closely together, suggesting that stereo processing already starts in area V1.

Since the seminal work of Sanger (1988) and Jenkin and Jepson (1988) exactly 25 years ago, the phase model has attracted a lot of attention. Being a very intuitive and appealing model, its simplicity seemingly contradicts that lot of attention. Indeed, results obtained with real images, also with random dot stereograms, are not very good, and one might say that the model appears to be a blessing, but a cursed one. Most researchers may be aware of the model's sting, but seem to have problems in locating and characterising that sting. Part of the problem may be due to exaggerated expectations: a very simple model which should satisfy two conflicting requirements, namely to provide a reconstruction of smooth and continuous surfaces, yet in combination with sharp and precisely located depth transitions.

In this theoretical paper we will therefore focus on the localisation problem of the simple phase model. For an overview of alternative biological disparity models see (Read and Cumming, 2007). In the next two sections we present a brief overview of earlier work, including Sanger's, to illustrate the "sting," the constraints which have been and still are applied in estimating disparity, and some important assumptions which should be taken into account. Section 4 deals with the simplest case, namely positive lines. Negative lines and edges are analysed in Section 5. In Section 6 we analyse two combinations: line clusters and bars. In Section 7 we present a summary of the phase model and a way to circumvent the problems. How local phase should be used in one-sided views is described in Section 8, whereas cyclopean vision is described in Section 9. An alternative phase model which is based on complex conjugate responses is analysed in Section 10. We conclude with a small discussion in Section 11.

2 MANY CONSTRAINTS AND YET POOR RESULTS

Like others, Sanger may have realised the localisation problem, but he focused on random dot stereograms (RDSs). He introduced a set of constraints together with a smoothing step in order to obtain model predictions which, at least, resemble perceived disparity of the RDSs. First, Gabor filter responses are thresholded such that those with insufficient amplitude and therefore meaningless phase are ignored.

M. H. du Buf J., Terzic K. and M. F. Rodrigues J..
Phase-differencing in Stereo Vision - Solving the Localisation Problem.
DOI: 10.5220/0004349702540263
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing* (BIOSIGNALS-2013), pages 254-263
ISBN: 978-989-8565-36-5

Second, a confidence is computed using responses of the left and right views, with similar responses yielding a higher confidence, and the disparity estimates obtained at multiple filter scales are averaged by using the confidences. This is supposed to suppress disparity estimates distorted by noise in each filter's frequency band as well as those beyond the detection limit of each filter. The above criteria are all merely based on amplitudes, not on any phase or disparity information. Third, a second-level confidence is used to exclude scales with disparity estimates which deviate from the (weighted) average, i.e., outliers, now taking into account both the amplitudes and disparity estimates at all applied scales. Finally, the second-level confidence is used to create a smooth surface, assuming that nearby points should have similar disparities. This is a nonlinear spatial filtering because of the weighting of neighbouring disparity values by their confidences. Despite all processing before and after computing the phase differences, with emphasis on image noise, detection limits and local consistency of disparity estimates, Sanger's results are extremely noisy and imprecise in terms of localisation; see also (Jenkin and Jepson, 1988).

The same observation holds when looking at more recent results. For example, Fröhlinghaus and Buhmann (1996) considered disparity estimation an ill-posed problem which requires regularisation in order to produce smooth results, and thereby lose, not to say sacrifice, good localisation. This effect is clearly visible in most if not all results which employ real images (Pauwels et al., 2012). Solari et al. (2001) presented an alternative phase-differencing model (see Section 10), perhaps more in line with biological processing, but without any postprocessing or regularisation. Their results are also very noisy, and neat vertical edges look crooked, as if they were distorted by some "random zig-zag filter."

Hence, one cannot deny that there must be some obscure problem lurking in phase differencing and, to the best of our knowledge, this "sting" has not yet been clearly identified. One of the most profound and recent analyses, by Monaco et al. (2008), addresses applicable constraints, but assumes as principal image model white noise, like in RDSs. The main constraints, as already formulated and applied in earlier work (Sanger, 1988; Fleet and Jepson, 1990; Fleet et al., 1991), serve to avoid regions where the instantaneous frequency (the derivative of the phase signal) deviates too much from the filter's central frequency, like at phase jumps caused by phase wrapping of the arctangent function, and to avoid singularities where the real and imaginary response components are zero and the phase is not defined. As we will see below, such constraints are not really beneficial, not to say completely useless, in the case of real images dominated by lines, bars and edges.

3 A VERY SIMPLE MODEL IN A VERY COMPLEX SYSTEM

In contrast to other approaches we will ignore RDSs and the white noise image model. Simple and complex cells in area V1 may code noise patterns, such as stochastic textures, but their main purpose is to code lines and edges for object recognition and brightness perception (Rodrigues and du Buf, 2009). When focusing on models for disparity, one easily forgets that there also are other cells for coding specific patterns, like individual bars and periodic gratings (du Buf, 2007) and dot patterns like the ones used in Sanger's RDSs (Kruizinga and Petkov, 2000). Our insight into the complexity of the visual system, with a myriad of different cell functions and a zillion of connexions, is still rather poor. It may make sense to develop models for specific purposes, like one model for RDSs and another one for lines and edges, before trying to develop a unified model.

We explicitly question the requirement that, at a very low level in the visual system such as in areas V1 and V2 etc., our visual system reconstructs a "replica" of our 3D environment. The idea of attributing disparity information to planar and curved surfaces may seem to make sense. It is the same idea as that behind brightness perception of surfaces: a realistic and precise reconstruction of scenes and objects, like a photograph, every pixel of the photograph even complemented by depth information. But philosophically that idea has a big problem. If 3D scenes are reconstructed at an early stage in vision, then there should be another, virtual observer in our brain who analyses that reconstruction. If that other observer also reconstructs, there should be yet another one. Hence, this reasoning leads to infinite regress.

Taking into account that the visual system applies hierarchical processing, from simple elementary features to increasingly more complex ones, from syntax to semantics, the only solution is to work with features. Like in brightness perception, where local brightness is only attributed to lines and edges in a multi-scale representation (Rodrigues and du Buf, 2009), the same can be done with disparity. Since many simple and complex cells in V1 (and other cells in higher areas) are disparity tuned, and since those cells provide an efficient vehicle for line and edge coding, disparity can be attibuted to detected lines and edges. This yields a sort of wireframe representation

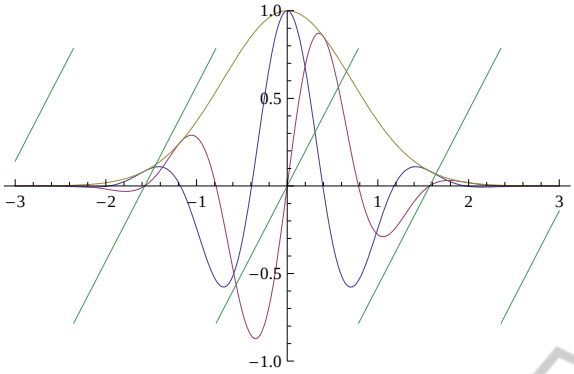

Figure 1: The example simple cells: real (cosine) and imaginary (sine) responses, the Gaussian envelope, and the wrapped phase divided by $\omega_0 = 4$. Note: $\pi/4 \approx 0.8$.

as used in computer graphics. Such a wireframe represents structural shape properties of objects, including existing patterns on their surfaces. The big advantage is that there is no localisation problem: if a line or edge is detected in the left view, the response (phase) in the right view at exactly the detected position can be employed. In other words, the localisation problem is reduced to a detection problem (Rodrigues and du Buf, 2009). Below we illustrate the localisation problem, and we show how phase information should be exploited.

4 THE POSITIVE LINE CASE

As usual, a complex Gabor function which is used to model simple and complex cells in V1 is defined by

$$G(x) = e^{-x^2/2\sigma^2 + i\omega_0 x} = e^{-x^2/2\sigma^2}(\cos\omega_0 x + i\sin\omega_0 x). \quad (1)$$

Below we will illustrate results using specific values: $2\sigma^2 = 1$ and $\omega_0 = 4$. The latter yields a period $T_0 = 1/F_0 = 2\pi/\omega_0 = \pi/2 \approx 1.6$. An input signal defined on x is convolved with $G(x)$, after which we can compute the real (Re) and imaginary (Im) responses of the even- and odd-symmetric simple cells. The amplitude (Mod) is $(\mathrm{Re}^2 + \mathrm{Im}^2)^{1/2}$ as a model of complex cells, and finally the phase θ is the argument (Arg) of the complex response. See Fig. 1.

The simplest case to study is one Dirac function with "amplitude" 1 at $x = 0$. Because of its sifting property, the response equals Eq. 1. The phase model yields $\theta(x) = \omega_0 x$, which comprises the filter's frequency ω_0. There are two possibilities to eliminate this frequency component from θ: (a) divide by ω_0, which implicitly assumes that some cells know this parameter, or (b) divide by the instantaneous frequency $\bar{\omega}(x) = \dot{\theta}(x)$, the derivative of $\theta(x)$. The use of the instantaneous frequency assumes that the cells do not have access to the value of parameter ω_0, but it requires an additional cell layer which implements the derivative operator. In this trivial case $\dot{\theta} = \omega_0$, such that $\theta/\dot{\theta} = x$. This means that the phase is an ideal vehicle for translation estimation, both for disparity and optic flow (Fleet and Jepson, 1990; Fleet et al., 1991).

The above analytic analysis yields a perfect result for translation estimation, even on $[-\infty, \infty]$ and for arbitrary values of σ and ω_0, but there is a big problem: Arg$G(x)$ requires the arctan operator, such that the phase is limited to $(-\pi, \pi]$, it is wrapped around, and it is periodic. Hence, only valid is $|\omega_0 x| \leq \pi$ such that $\theta/\dot{\theta} = x$ for $|x| < \pi/\omega_0 = T_0/2$, i.e., on exactly one period of $G(x)$. In the last formula we wrote $<$ instead of $\leq$ because of the derivative operator, which must avoid the two phase jumps of 2π, but this is of secondary importance.

The simplest case of disparity estimation involves one positive Dirac seen against a homogeneous background. We will focus on near disparity: a thin, white pole, for example that of a traffic sign, seen against a dark wall. Now we introduce left (L) and right (R) views with a shift of $\pm\delta$ relative to $x = 0$. Near disparity implies that in the left view the Dirac shifts to the right and in the right view it shifts to the left. The goal is to estimate the total shift of 2δ.

We have $L(x) = G(x - \delta)$ and $R(x) = G(x + \delta)$, therefore $\theta_R(x) = \omega_0(x + \delta)$, $\theta_L(x) = \omega_0(x - \delta)$, and let us define disparity as is usually done by $D(x) = \Delta\theta/\dot{\theta}$, where $\Delta\theta = \theta_R - \theta_L$. In this simple case we can divide $\Delta\theta$ by ω_0. Anyhow, the result is 2δ. This analytic solution is correct: it is the left-sided view in which the right-sided projection is shifted 2δ to the right. However, $D(x) = 2\delta$ holds for $-\infty < x < \infty$ because we are working with ideal and generalized functions. In practice we can limit the interval by thresholding the modulus (the responses of complex cells: the Gaussian envelopes of the Gabor filters), but only the maximum is well defined. Furthermore, the arctan operator must be applied, to both L and R. As a result, both θ_L and θ_R are restricted to $\pm\pi$. Normally, two constraints are applied (Monaco et al., 2008): if $\bar{\omega} = \dot{\theta}$, then $|\bar{\omega} - \omega_0| < \rho$, which serves to stay away from the phase discontinuities where the phase wraps around, also to avoid strong phase nonlinearities (see below), but here this is useless because $\bar{\omega} = \omega_0$, except at the phase jumps. The second constraint serves to avoid singularities where both real and imaginary responses are zero, which is also useless here.

What are the problems introduced by the arctan function? The two phases are linear and zero at the positions of the two Diracs, i.e., they are parallel and shifted functions. This means that there are

two crossings at $+\pi$ and two at $-\pi$, and we need to take the inner two for the validity range of $\Delta\theta$: this yields $\delta - \pi/\omega_0 < x < -\delta + \pi/\omega_0$, or $\delta - T_0/2 < x < -\delta + T_0/2$. But we do not know δ, we only obtain $D(x)$. A first conclusion is that the localisation of the Dirac is lost: $D(x) = 2\delta$ for $\delta - T_0/2 < x < -\delta + T_0/2$. Beyond these two limits, the phases are wrapped and $D(x)$ becomes negative: $D(x) = 2\delta - T_0$. And $D(x)$ is periodic (T_0). If $\delta = T_0/2$, perfect localisation is obtained (2δ only in $x = 0$), but this is flanked by negative values over the entire period. If δ increases, or ω_0 decreases, the value of $D(x \approx 0)$ even becomes negative! One can easily see that for $2\delta \to 0$ almost the entire period T_0 is correct ($D = 2\delta$) but localisation is completely lost. Therefore, a second conclusion is that the disparity estimates may be correct on some intervals, but wrong (negative) on the remaining intervals. And these intervals are a function of scale, i.e., of ω_0 relative to the real disparity. If multiple scales are applied, the problem can only be solved by combining the various results, like Sanger did, but, given his noisy results, this is not an easy problem. As we will see in the next section, the confusion is even much worse.

5 NEGATIVE LINES AND EDGES

A dark line against a homogeneous background, for example a thin black pole in front of a white or gray wall, is of course modelled by a negative Dirac function, and intuitively one expects results similar to a positive Dirac but with negated responses. Indeed, we obtain the following response:

$$\begin{aligned} F(x) &= -G(x) = e^{-x^2/2\sigma^2 + i(\omega_0 x + \pi)} \\ &= e^{-x^2/2\sigma^2}(-\cos\omega_0 x - i\sin\omega_0 x). \end{aligned} \quad (2)$$

The phase is $\theta(x) = \omega_0 x + \pi$, which means that in practice there is a phase discontinuity exactly at the line's position. Therefore, translation estimation is *not* similar to that in case of a positive Dirac: (1) at $x = 0$ the phase is undefined or may switch randomly between π and $-\pi$ because of noise; (2) for $x > 0$ the relative translation is $-\pi + \omega_0 x$, so one could add π to correct this; (3) for $x < 0$ it is $\pi + \omega_0 x$ and one could subtract π.

In the case of disparity (two shifted, negative Diracs) we have $\theta_R = \omega_0(x + \delta) + \pi$ and $\theta_L = \omega_0(x - \delta) + \pi$, such that $\Delta\theta = 2\delta\omega_0$, but this is the analytical solution for $x \in [-\infty, \infty]$. The difference of the *wrapped* phases is *negative* around $x = 0$ for small δ due to the two discontinuities at $x = \pm\delta$. To be precise, for $\delta < \pi/\omega_0 = T_0/2$ the disparity estimate at

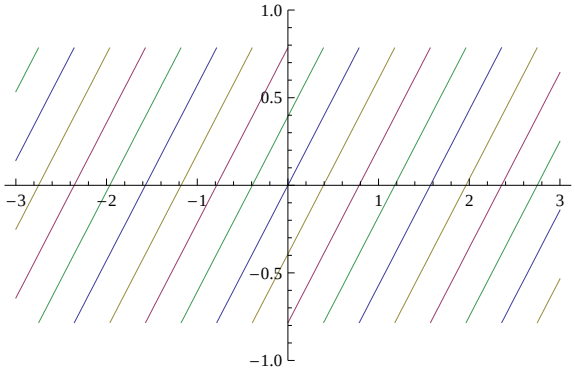

Figure 2: Phase plots (divided by $\omega_0 = 4$) of responses to positive and negative lines and edges at $x = 0$. Crossing the vertical axis at $x = 0$ are, from top to bottom: L^-, E^-, L^+, E^+ and L^-.

$x = 0$ is $2\delta - 4\pi/\omega_0 = 2\delta - 2T_0$ because of the two phase jumps of 2π, and for $\delta > T_0/2$ it is $2\delta - T_0$. In other words, at $\delta = T_0/2$ the disparity estimate jumps from one wrong value to another wrong value. The phase constraint $|\bar{\omega} - \omega_0| < \rho$ can be used to skip the phase discontinuities, but these are only a few points on x and all other disparity estimates on x are completely useless.

Edges are much more complicated to deal with analytically, because responses of complex Gabor filters are complex error functions. Fortunately, it has been shown that complex error functions can be well approximated by Gabor functions, by tweaking the two parameters a little bit (du Buf, 1993). He called this an abnormal scaling of the Gabor function. If a normal scaling is obtained by σs and ω_0/s, with s the scale parameter, an abnormal scaling, i.e., approximation of the error function which is the response of a normally scaled Gabor filter, is obtained by $\bar{\sigma} = \sigma/a$ and $\bar{\omega}_0 = \omega_0/b$, where σ and ω_0 are the parameters of the Gabor filter. Below we apply this approximation and deal with edge responses through Gabor functions. We will not include the proportionality constant of the error function, because this appears in the modulus of the Gabor response and it simply disappears in the phase. For the sake of simplicity we even write σ and ω_0 in stead of $\bar{\sigma}$ and $\bar{\omega}_0$ (please recall that the instantaneous frequency must be used, which is approximately $\bar{\omega}_0$).

Going from left to right, a positive edge E^+ is defined as a positive transition in $x = 0$: a step with value -1 for $x < 0$ and $+1$ for $x > 0$. Similarly, a negative edge is defined by $E^- = -E^+$. The response to a positive edge is described by

$$\begin{aligned} F(x) &= G^+(x) = e^{-x^2/2\sigma^2 + i(\omega_0 x - \pi/2)} \\ &= e^{-x^2/2\sigma^2}(\sin\omega_0 x - i\cos\omega_0 x). \end{aligned} \quad (3)$$

The phase is $\omega_0 x - \pi/2$, i.e., in $x = 0$ it is $-\pi/2$ (we note that this is different from the value $\pi/2$ obtained by du Buf (1993), because he used correlation whereas here we use convolution). Similarly, the response to a negative edge is described by

$$\begin{aligned} F(x) &= G^-(x) = e^{-x^2/2\sigma^2 + i(\omega_0 x + \pi/2)} \\ &= e^{-x^2/2\sigma^2}(-\sin\omega_0 x + i\cos\omega_0 x). \end{aligned} \quad (4)$$

The phase is $\omega_0 x + \pi/2$, i.e., in $x = 0$ it is $\pi/2$. The phase values at the edges imply that the pivot points through which the linear phase plots rotate as a function of the scale (ω_0) are at $x = 0$ and $\theta = \pm\pi/2$, with the usual discontinuities at $\pm\pi$. Figure 2 summarises the different phase behaviours of positive and negative lines and edges at position $x = 0$.

Suppose that there is only one positive edge in a neighbourhood with a L-R shift of 2δ. We therefore have

$$L(x) = G^+(x-\delta) = e^{-(x-\delta)^2/2\sigma^2 + i(\omega_0(x-\delta) - \pi/2)}. \quad (5)$$

The phase is $\theta_L = \omega_0(x-\delta) - \pi/2$. We also have

$$R(x) = G^+(x+\delta) = e^{-(x+\delta)^2/2\sigma^2 + i(\omega_0(x+\delta) - \pi/2)} \quad (6)$$

with a phase $\theta_R = \omega_0(x+\delta) - \pi/2$, such that $\Delta\theta = \theta_R - \theta_L = 2\omega_0\delta$ and $\dot{\theta}_L = \dot{\theta}_R = \omega_0$. However, phase wrapping implies that $\Delta\theta$ may be negative around $x = -\delta$ and positive around $x = \delta$, obviously with a discontinuity in between, and this discontinuity shifts as a function of the scale (ω_0). For example, if $\delta = T_0/4$, the discontinuity will be exactly at $x = 0$. To the right, the phase difference is correct ($\omega_0 2\delta$) until θ_R reaches π, i.e., for $0 < x < T_0/2$. To the left, the phase difference is wrong ($\omega_0 2\delta - 2\pi$) because θ_L wraps around at $x = 0$. This phase jump of 2π in combination with $\Delta\theta = \theta_R - \theta_L$ divided by ω_0 yields the wrong disparity $2\delta - T_0$. The wrong phase difference and disparity are obtained until θ_R reaches $-\pi$, i.e., for $-T_0/2 < x < 0$. Hence, one half of the period is correct and the other half is wrong, this is repeated periodically (T_0), and there is absolutely no localisation except if we truncate the result somehow using the modulus (responses of complex cells). Disparity estimation in the case of one negative edge is similar, yet with the same problems.

6 LINE CLUSTERS AND BARS

If lines and edges are well separated, they can be detected individually because the modulus will show distinct maxima and their phases will not be distorted by interference effects (du Buf, 1993). The question is what happens when they are closer together, forming clusters of lines and bars. We first analyse line clusters and then bars.

There are many possibilities concerning distances and amplitudes. For the sake of simplicity let us take three equidistant lines (Diracs) with "amplitude" 1 and centered around $x = 0$:

$$F(x) = G(x) + G(x+\delta) + G(x-\delta). \quad (7)$$

What is the influence of the Diracs at $x = \pm\delta$ on the response of the Dirac at $x = 0$? The exact solution is

$$F(x) = G(x) \cdot [1 + 2e^{-\delta^2/2\sigma^2}\cos(\omega_0\delta + i\frac{2x\delta}{2\sigma^2})], \quad (8)$$

where

$$2\cos(\omega_0\delta + i\frac{2x\delta}{2\sigma^2}) = e^{-2x\delta/2\sigma^2 + i\omega_0\delta} + e^{2x\delta/2\sigma^2 - i\omega_0\delta}. \quad (9)$$

Using a Taylor series around $x = 0$ (or $e^z \approx 1 + z$) results in

$$F(x)/G(x) = 1 + 2e^{-\delta^2/2\sigma^2}(\cos\omega_0\delta - i\frac{x\delta}{\sigma^2}\sin\omega_0\delta). \quad (10)$$

In order not to influence the phase of the response of $G(x)$ around $x = 0$, i.e., $\omega_0 x$, this function must be real. We see that there are two possibilities. One is $\delta \ll \sigma^2$, which relates the distance to the width of the envelope. The other is $\sin\omega_0\delta = 0$. This means $\omega_0\delta = k\pi$, or $\delta = k\pi/\omega_0 = kT_0/2$. There are only two possibilities in practice: 0 and π/ω_0. When $\delta = 0$, (8) results in $F(x) = 3G(x)$, as expected: three Diracs in $x = 0$. When $\delta = \pi/\omega_0 = T_0/2$, (8) results in

$$F(x)/G(x) = 1 - e^{\pi/\sigma^2\omega_0}(e^{-x} + e^{x})e^{-\pi^2/\sigma^2\omega_0^2}, \quad (11)$$

which is also real. In contrast to these two cases, we can expect the biggest problem (worst case?) when in (10) the imaginary part is maximum: $\sin\omega_0\delta = 1$ or $\delta = \pi/2\omega_0 = T_0/4$. Substituting this δ in (8) and approximating $\exp(-x) - \exp(x)$ by $-2x$ gives

$$F(x)/G(x) = 1 - ix2e^{\pi/2\sigma^2\omega_0 - \pi^2/8\sigma^2\omega_0^2} = 1 - ixA. \quad (12)$$

The modulus is $(1 + x^2A^2)^{1/2}$ and the phase is $-\arctan(xA)$. With $\arctan(xA) \approx xA$ for $|xA| \leq 1$, i.e., $x \approx 0$, we then obtain

$$F(x) = e^{-x^2/2\sigma^2}\sqrt{1 + x^2A^2}e^{i(\omega_0 x - Ax)}. \quad (13)$$

When dividing $\theta(x) = (\omega_0 - A)x$ by ω_0, translation estimation gives $\theta(x)/\omega_0 = (\omega_0 - A)x/\omega_0 = x(1 - A/\omega_0)$. Substitution of the values of σ and ω_0 gives $A = 3.76$, $A/\omega_0 = 0.94$, and $x(1 - A/\omega_0) = 0.06x$. Obviously, this is not correct, and wrong translation estimation will result in wrong disparity estimation. When using the instantaneous frequency instead, we

divide $\theta(x) = (\omega_0 - A)x$ by $\dot{\theta}(x) = \omega_0 - A$ and obtain $\theta/\dot{\theta} = x$, which is the correct translation estimation. However, we forgot several approximations like $|x| \approx 0$ and $|xA| \leq 1$. The latter means $|x| \leq 1/A \approx 0.25$, which is of the order of $T_0/6$. Hence, at and beyond $T_0/6$ the analytical result is not reliable. Please recall that we analysed the case $\delta = T_0/4$. Hence, we saw that the instantaneous frequency must really be used and we can expect that a small cluster of close lines may behave as one line.

Concerning bars, a positive bar with amplitude 1, centered at $x = 0$ and with width 2δ, is defined by $B^+(x) = 0.5\{E^+(x+\delta) + E^-(x-\delta)\}$. If δ is very small, we expect responses similar to those of lines (Diracs), but because there are two edges the analysis is more complicated. Using the edge responses defined before, we obtain

$$\begin{aligned} R(x) &= \frac{1}{2}\{G^+(x+\delta) + G^-(x-\delta)\} \\ &= \frac{1}{2}\{e^{-(x+\delta)^2/2\sigma^2}(-ie^{i\omega_0(x+\delta)}) \\ &\quad + e^{-(x-\delta)^2/2\sigma^2}(ie^{i\omega_0(x-\delta)})\}. \end{aligned} \quad (14)$$

In order to see what this means, we can first look at the response in the centre:

$$\begin{aligned} R(0) &= \frac{1}{2}e^{-\delta^2/2\sigma^2}\{-ie^{i\omega_0\delta} + ie^{-i\omega_0\delta}\} \\ &= \sin\omega_0\delta \cdot e^{-\delta^2/2\sigma^2}. \end{aligned} \quad (15)$$

This is real and we can distinguish the following cases based on the sine function (please recall that the width is 2δ and that the period of the Gabor function is $T_0 = 2\pi/\omega_0$):

- For $\delta = 0$ there is a singularity in $x = 0$ because both the real and imaginary responses are zero. One could take the limit $\delta \to 0$ to approximate the Dirac case, but the width approaches 0 whereas the bar's amplitude remains 1.
- If $0 < \omega_0\delta < \pi$, i.e., $2\delta < T_0$, the phase is zero in $x = 0$, which corresponds to that of a positive line. Here we can have a relatively narrow bar which behaves as a line, but also a broad one because $0 < 2\delta < T_0$, i.e., a width up to one period. In the case of a broad bar, the two edges could be detected separately.
- For $\omega_0\delta = \pi$ ($2\delta = T_0$) there is again a singularity at $x = 0$.
- For $\pi < \omega_0\delta < 2\pi$ the phase in $x = 0$ is $\pm\pi$, which indicates a negative line, but this occurs for $T_0 < 2\delta < 2T_0$ where the individual edges are detected and the modulus in $x = 0$ is small.

This was the response in the centre, so now we can analyse the responses at the edge positions. If we

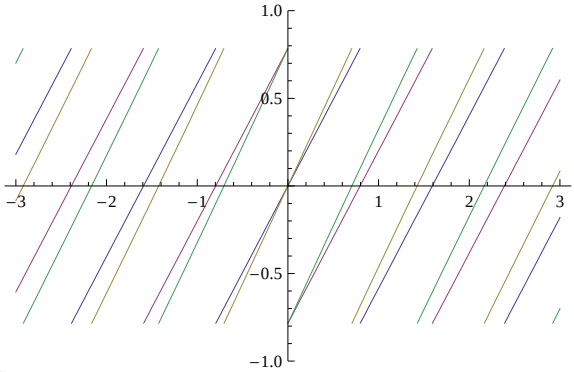

Figure 3: Phase plots of three lines at positions -0.1, 0 and 0.1 and bars with edges at -0.1 and 0.1. A cluster of three close positive lines and a narrow positive bar behave as one positive line (phase through the origin). The negative versions behave as one negative line (the phase crosses the vertical axis at ± 0.8). Note: $\pi/4 \approx 0.8$.

choose as example $\omega_0\delta = \pi/2$ or $2\delta = T_0/2$ (see the second case above), it follows from (14) that

$$R(-\delta) = -\frac{1}{2}i(1 + e^{-2\delta^2/\sigma^2}). \quad (16)$$

Hence, $R(-\delta) \approx -i$ if $\delta^2 \ll \sigma^2$ (the phase is $-\pi/2$). Similarly, we obtain $R(\delta) \approx i$ (the phase at the other edge is $\pi/2$). Since the response $R(0)$ is real, the phase can be approximated by $\theta(x) \approx (x/\delta)(\pi/2) = \omega_0 x$ inside the bar. Outside the bar, the response and therefore the phase is dominated by the positive edge on the left side, and by the negative edge on the right side, both with a component $\omega_0 x$ (because of the approximation of the complex error function this is $\bar{\omega}_0 x$) and with a phase of $\pm\pi/2$ at the two edges.

Figure 3 shows that a cluster of close lines and a narrow bar indeed behave as a line, both in the positive and negative cases. Because we divided the phase by ω_0 in stead of the instantaneous frequency, the slopes are slightly different.

7 STINGS AND TWEEZERS

From the above analyses we may conclude that the simple phase model is not really simple. There are serious problems involved in obtaining reliable disparity estimates, and these can explain the poor results which have been obtained in previous work (Sanger, 1988; Jenkin and Jepson, 1988; Fröhlinghaus and Buhmann, 1996; Solari et al., 2001; Pauwels et al., 2012). And instead of one "sting" of the model we have identified three stings: (1) phases and phase differences are not localised, which may be an advantage when creating a smooth depth map but it comes at the cost of sacrificing any neat depth transitions;

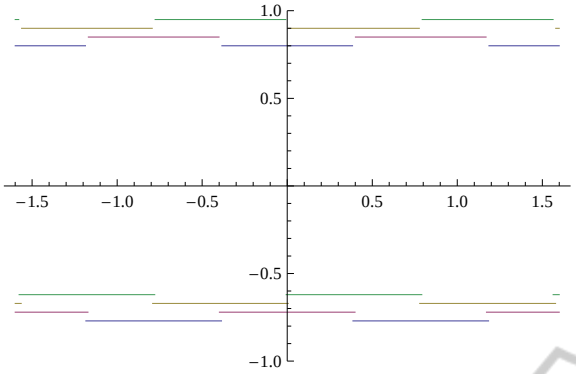

Figure 4: Plots of $\Delta\theta/\omega_0$ in the case of $\delta = T_0/4 \approx 0.4$, hence a real disparity of 0.8. In this case half the period is correct (at top) and the other half is wrong (at bottom). In order to avoid clutter, the disparities of a negative line, positive edge and negative edge are shifted higher 0.1, 0.2 and 0.3, respectively, relative to that of a positive line. If δ increases, the intervals with the correct disparity (at top) become smaller, those with wrong disparity (at bottom) become larger.

(2) phase differences may be correct on some intervals of the filter's period, but they are wrong on the remaining intervals, and these intervals are repeated periodically; and (3) the intervals with correct and wrong phase differences depend on the wavelengths of the filters in relation to the real disparity, but also on the local image structure, where negative lines are even disastrous. In other words, the phase difference may be easy to compute but yields totally confusing and even contradictory results. Figure 4 summarises the confusion. If a fronto-parallel planar surface with a certain disparity is decorated with periodic and well separated positive and negative lines and edges, disparity estimation is doomed to produce a randomly corrugated surface. In random dot stereograms (Sanger, 1988; Jenkin and Jepson, 1988) and in white noise stereograms (Monaco et al., 2008; Solari et al., 2001) there are always local fluctuations which cause random phases, so how can one expect to obtain e.g. flat surfaces? It seems that all processing applied targets smooth surfaces. Even more advanced processing schemes, in which disparity estimation starts at a coarse scale and is refined at progressively finer scales, seem to target only smooth surfaces such that well-defined depth transitions at high-contrast object boundaries often produce disastrous results (Pauwels et al., 2012).

Focusing on low-level processing, also a solution became obvious for tweezing the stings from the model: the only way to obtain reliable and precise disparity information is to check the local image structure, i.e., to detect lines and edges and to attribute disparity only to detected events. Therefore one solution is to combine the phase with line and edge detection: we detect a line or an edge at a certain position in L, and take the phase in R at the same position. We will see that we need to detect the same event type in L and R in order to avoid problems. One might say that we therefore apply feature matching: a positive edge in R must match a positive edge in L, hence the spatial distance is available and therefore explicitly the disparity. However, we assume that the cells in the retinotopic mapping do not have access to their absolute position. We only assume that there are cell clusters in the direction of the simple cells and in a certain neighbourhood with a size related to the scale (and event type; see below). This is explained in the following section. The next section deals with cyclopean vision.

8 USING THE PHASE OF ONE VIEW

In what follows we assume that the system detects events in L and analyses R for making sure that there is no other near event which destroys the correct phase information. Now, assume that event detection in L and R works, i.e., there is a good detection model. As before, we assume near disparity of 2δ with a right-shift of δ in L and a left-shift of δ in R. Hence, events are detected in L at $x = \delta$. In R we have

$$R(x) = e^{-(x+\delta)^2/2\sigma^2 + i(\omega_0(x+\delta)+\phi_0)} \tag{17}$$

and, similarly, in L we have

$$L(x) = e^{-(x-\delta)^2/2\sigma^2 + i(\omega_0(x-\delta)+\phi_0)} \tag{18}$$

with phases $\theta_R = \omega_0(x+\delta)+\phi_0$ and $\theta_L = \omega_0(x-\delta)+\phi_0$. The constant phase component ϕ_0 is 0 for a positive line (L^+), $\pm\pi$ for a negative line (L^-), $-\pi/2$ for a positive edge (E^+), and $\pi/2$ for a negative edge (E^-). In all four cases we substitute $x = \delta$ in both $L(x)$ and $R(x)$ and obtain a disparity $\Delta\theta/\dot{\theta} = 2\delta$ in $x = \delta$ because $\dot{\theta} = \dot{\theta}_L = \dot{\theta}_R = (\dot{\theta}_R + \dot{\theta}_L)/2 = \omega_0$.

What are the constraints? Because there are no singularities in case of isolated events, we can restrict the analysis to phase discontinuities, normally expressed by the constraint $|\dot{\theta} - \omega_0| < \rho$. In our case, where we detect events in L with specific phase values, there are no phase jumps at event positions (only negative lines must be treated carefully) and we must only avoid phase jumps in R. If we increase δ starting with a small value, the phase difference increases linearly until θ_R reaches π, where it wraps around to $-\pi$, and then increases again linearly. Hence, the above constraint only serves to avoid the phase jump, but not the region beyond it. If the primary phase region

is the one before the first phase jump, then which primary regions yield a valid phase difference? If we make phase plots with the different values of ϕ_0, we can easily see that the translation before θ_R reaches π is largest for a negative line and smallest for a negative edge. The specific ranges are: $2\delta_{L^+} < (1/2)T_0$, $2\delta_{L^-} < T_0$, $2\delta_{E^+} < (3/4)T_0$, and $2\delta_{E^-} < (1/4)T_0$. Please recall that in case of edges we should have written $\bar{T}_0$ because of $\bar{\omega}_0$ which results from approximating the complex error function responses by complex Gabor functions, and this is obtained by computing the instantaneous frequency $\dot{\theta} \approx \bar{\omega}_0$.

Hence, there are four different valid intervals for the four event types. Beyond these intervals, i.e., for larger disparities, in all four cases we get $2\delta - T_0$ because of the phase jump of -2π, which, divided by $\dot{\theta}$ or ω_0, results in $-T_0$. This phase difference is negative and therefore completely wrong. Although the constraint $|\dot{\theta} - \omega_0| < \rho$ may be met at all positions except at the phase jump, which can be skipped, this is not sufficient because both (1) the phase jump in R must be skipped and (2) the region with the negative disparity estimate must be avoided. We assume of course that the cells do not have access to the value of T_0 and that no circuit for phase unwrapping exists. The above analysis only implies that there exists a circuit which checks events in a neighbourhood with event-related size. Above results are valid for isolated events, i.e., if there is only one in a receptive field, and at all possible scales (the T_0 parameter). In case of dense patterns of lines and edges, the smallest scale will be the limiting factor unless specific circuits tuned to such patterns are assumed: periodic grating cells (du Buf, 2007) and dot pattern cells (Kruizinga and Petkov, 2000).

9 CYCLOPEAN VISION

Above we assumed that disparity is attributed to the left view. A similar analysis must be applied in case of attributing disparity to the right view. However, these attributions are only common practice because the ground truth for verifying or validating results is often only available for the left or right view. Apart from binocular rivalry in specific cases, we do not experience the world in terms of left or right views. Hence, let us therefore focus on cyclopean vision, still considering near disparity. In this case we must assume that left- and right-shifted events are detected in respectively R *and* L, and that there is a symmetry analysis by means of symmetric cell clusters or cells with symmetric dendritic fields. Hence, we apply symmetric event detection and compute the phase difference in $x = 0$, i.e., the centre of a neighbourhood.

Because of the symmetry, detection limits will be different because the phases of both the left and right projections may wrap around. In case of lines, both will wrap around, but in case of edges only one will do. Because of the special responses in case of negative lines, which results in a negative disparity estimate for small δ, a correction must be applied after computing the phases. Since $\theta_R = \omega_0(x+\delta) - \pi$ to the right of the discontinuity, and $\theta_L = \omega_0(x-\delta) + \pi$ to the left of the discontinuity, we must correct them by adding π to θ_R and subtracting π from θ_L, such that $\Delta\theta = 2\omega_0\delta$. The specific ranges are then: $2\delta_{L^+} = 2\delta_{L^-} < T_0$, and $2\delta_{E^+} = 2\delta_{E^-} < (1/2)T_0$, again with $\bar{T}_0$ in case of the edges. Beyond these limits, the disparity in $x = 0$ will be $2\delta - 2T_0$ in case of lines (two wrapping phases, i.e., twice 2π), and $2\delta - T_0$ in case of edges (one wrapping phase or 2π). Hence, in order to obtain valid phase differences the above detection intervals must be respected. The symmetrically detected events are attributed to the centre of the detection interval ($x = 0$), and the phase difference there is used, yielding a cyclopean feature map. The advantage of a cyclopean model is that both left and right intervals must be checked for symmetric events. This way we can detect the presence of other nearby events, both on the left side and on the right side of the symmetric events, which could influence the correct phases. This may solve the noise problem, but cannot solve occlusions where detail is visible in one view but not in the other view.

10 COMPLEX CONJUGATE RESPONSES

There are two ways to compute the phase difference. The first way, as in the simple phase model treated above, is to extract the left and right phases separately and to take the difference. In R we have

$$R(x) = e^{-(x+\delta)^2/2\sigma^2 + i(\omega_0(x+\delta)+\phi_0)} \qquad (19)$$

and, similarly, in L we have

$$L(x) = e^{-(x-\delta)^2/2\sigma^2 + i(\omega_0(x-\delta)+\phi_0)} \qquad (20)$$

with phases $\theta_R = \omega_0(x+\delta) + \phi_0$ and $\theta_L = \omega_0(x-\delta) + \phi_0$. The constant phase component ϕ_0 is different for the four event types, but it is the same in L and R, hence $\Delta\theta = \omega_0 2\delta$. However, $\mathrm{Arg}R(x)$ and $\mathrm{Arg}L(x)$ implies that the arctan function must be applied twice, and both θ_R and θ_L can be wrapped.

The second way is to multiply $R(x)$ and $L(x)$ because the phase exponentials in (19) and (20) will add

up. However, because we want the phase difference instead of the sum, we need to use one complex conjugate ($*$) response:

$$\begin{aligned} P(x) &= R(x)L^*(x) \\ &= e^{-(x+\delta)^2/2\sigma^2} e^{-(x-\delta)^2/2\sigma^2} \\ &\quad \times e^{i(\omega_0(x+\delta)+\phi_0)} e^{-i(\omega_0(x-\delta)+\phi_0)}, \quad (21) \end{aligned}$$

such that $\mathrm{Arg}P(x) = \Delta\theta = \omega_0 2\delta$ (Fleet and Jepson, 1990; Solari et al., 2001). In this case the arctan function needs to be applied only once, so the resulting phase difference may still be wrapped. This means that a correct phase difference will be obtained for $\omega_0 2\delta < \pi$, or $2\delta < \pi/\omega_0 = T_0/2$. This is the only criterion applicable to all four event types! For bigger disparities, the phase jump of -2π, after division by ω_0, yields the wrong value $2\delta - T_0$, or even $2\delta - kT_0$ with $k = 1, 2, \ldots$. In addition, both the correct and wrong disparity estimates are obtained for $x \in [-\infty, \infty]$. In contrast to the case of computing the phase difference by subtraction, here there are no subintervals with correct and wrong phase differences, either it is correct or it is wrong on $[-\infty, \infty]$, i.e., within the entire receptive field. Hence, localisation is even worse, and this may explain the very poor results obtained with this model (Solari et al., 2001).

The above solutions are based on the analytical expressions (19), (20) and (21). In practice, we have at each position (and each scale) the responses of the even- and odd-symmetric simple cells for computing the phase angles. In case of Eq. (21) this means that, if $R = (a+ib)$ and $L = (c+id)$, $RL^* = (a+ib)(c-id) = (ac+bd)+i(bc-ad)$, i.e., a simple combination of the addition and subtraction of multiplied components before applying the arctan function.

In the case of both solutions, the resulting phase difference must be divided by ω_0. If the individual phases of the left and right views are computed, one can take the derivative of the phase signal, the instantaneous frequency, because $\dot{\theta} = \omega_0$. In practice the derivative may deviate from ω_0 because of response interference effects, and these effects will occur at shifted positions if the disparity is not zero. Therefore one can opt for different solutions, like $(\theta_R/\dot{\theta}_R - \theta_L/\dot{\theta}_L)$ or $2(\theta_R - \theta_L)/(\dot{\theta}_R + \dot{\theta}_L)$ at any position x.

In the case of the second solution based on the complex conjugate response, Eq. 21, the individual phases are not computed, hence they cannot be differentiated. One can use the following solution (Fleet and Jepson, 1990; Solari et al., 2001). If $R(x) = C(x) + iS(x)$, where C and S are the cosine and sine components of the local response (responses of the even- and odd-symmetric simple cells), and if the modulus (response of complex cells) is $\rho = (C^2 + S^2)^{1/2}$, then $\mathrm{Im}(R^*\dot{R})/\rho^2 = \omega_0$ analytically, and this can be computed by using $(C\dot{S} - \dot{C}S)/(C^2+S^2)$. This implies that two derivatives must be computed, and this can be done for R_R and R_L, hence four derivatives can be involved. This solution is rather mathematical, so could it also be biological? As for now, no answer is readily available.

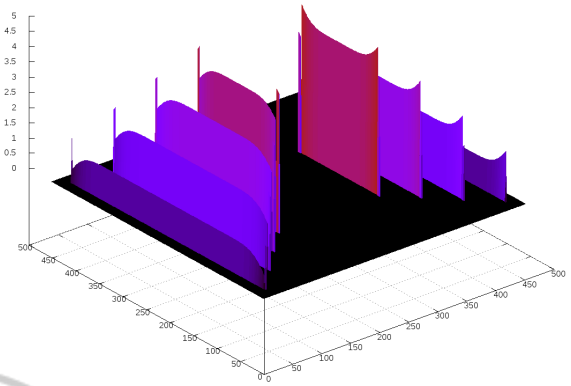

Figure 5: The 3D wireframe representation obtained by using a discrete pyramid, consisting of a few line squares, with increasing disparity at the top layers. Only filters tuned to vertical lines were employed. The representation has been rotated artificially in order to show the structure. See text for an explanation.

11 DISCUSSION

Suppose we do want to use phase information, assuming, like Sanger (1988) did, that the simple cells code the phase implicitly. However, we saw that the model applied by Sanger, namely to compute the average of the phase differences of all frequency bands, even when these are weighted by some combination of the responses of complex cells (the moduli), cannot be adequate. In addition, the usual constraints for avoiding phase jumps and singularities are not sufficient, even useless.

In this paper we have analysed a few responses, although this work could continue by analysing response interference effects in order to develop the most precise disparity model, but still the phase-differencing one. As we proposed, disparity estimation should be closely connected to line and edge detection. We tested this integration by using a synthetic stereo image which comprises lines, forming a square, discrete pyramid in wich the top layers have a larger disparity. We only applied filters tuned to vertical lines, hence only vertical lines are detected and complemented by disparity estimates. The resulting 3D wireframe representation is shown in Fig. 5. It captures all structural information of the pyramid, but no

surfaces. The horizontal lines of the pyramid can also be detected, perhaps with a residual disparity component, but only if the exact geometry of the projections in our eyes is taken into account (Read and Cumming, 2007). The spikes and curved parts at the line ends in Fig. 5 are caused by the fact that there are also horizontal lines in the vertically-aligned receptive fields, and only in the left or right half of the fields. These "half lines" cause different responses on the left and right side of the pyramid, because they are on the positive and negative half of the sinusoidal component of the Gabor filters, thereby influencing the phase near and at the corners of the squares. This complication can be explained theoretically, and it cannot be solved without assuming more advanced processing at a higher level than area V1. Hence, we cannot expect that all problems of phase differencing can be solved at a very low level of the visual system.

The virtual 3D wireframe representation captures all structural information of the pyramid, but no surfaces, unless the surfaces are textured. So how does our visual system manage to create continuous surfaces when they are *not* textured? This question is related to local feature and depth integration at a low level, as we did here, but also to learned and global object interpretation at a high level, likely the result of experience in combining visual with haptic (tactile) information in early childhood. Furthermore, there may exist some "filling-in" processes, for example to "hide" the blind spots of the retinae, but these occur at a very high level (O'Regan, 1998).

The necessary circuitry in case of cyclopean vision is very limited. First, there are circuits which detect events on the basis of simple and complex cells, in both the left and right views (Rodrigues and du Buf, 2009). Second, a level of gating cells with symmetric dendritic fields analyses local neighbourhoods: at the furthest points (the two longest dendrites with a length corresponding to the valid disparity range divided by two) they receive excitatory input if identical events are detected there; the other dendrites in between receive inhibitory input if asymmetric events are detected. A gating cell only passes the output of a third cell complex, which extracts the phases from the simple cells, their derivatives, and the phase difference. The gating cell complex also codes the type of symmetrically detected events at position $x = 0$ for obtaining a cyclopean representation. As a result, disparity is attributed to detected lines and edges with one, "centralised" view, in a way as used in the modeling of solid objects in computer graphics: the wireframe representation. As mentioned before, it does not make sense to reconstruct 3D objects with all entire surfaces at an early stage in vision, because our visual system applies a hierarchical processing strategy and the goal is to obtain a symbolic, semantic representation.

ACKNOWLEDGEMENTS

This work was supported by the Portuguese Foundation for Science and Technology (pluri-annual funding of LARSyS) and EU project NeuralDynamics FP7-ICT-2009-6, PN: 270247.

REFERENCES

du Buf, J. (1993). Responses of simple cells: events, interferences, and ambiguities. *Biol. Cybern.*, 68:321–333.

du Buf, J. (2007). Improved grating and bar cell models in cortical area V1 and texture coding. *Image Vision Comput.*, 25(6):873–882.

Fleet, D. and Jepson, A. (1990). Computation of component image velocity from local phase information. *Int. J. Comput. Vision*, 5(1):77–104.

Fleet, D., Jepson, A., and Jenkin, M. (1991). Phase-based disparity measurement. *CVGIP: Image Understanding*, 53(2):198–210.

Fröhlinghaus, T. and Buhmann, J. (1996). Regularizing phase-based stereo. In *Proc. of ICPR*, pages 451–455.

Jenkin, M. and Jepson, A. (1988). The measurement of binocular disparity. *Computational Processes in Human Vision: An interdisciplinary perspective, ed. Z. Pylyshyn, Ablex Press, Norwood, NJ*, pages 69–98.

Kruizinga, P. and Petkov, N. (2000). Computational model of dot-pattern selective cells. *Biol. Cybern.*, 83(4):313–325.

Monaco, J., Bovik, A., and Cormack, L. (2008). Nonlinearities in stereoscopic phase differencing. *IEEE Trans. on Image Processing*, 17(9):1672–84.

O'Regan, J. K. (1998). No evidence for neural filling-in - vision as an illusion - pinning down "enaction". *Behavioral and Brain Sciences*, 21(6):767–768.

Pauwels, K., Tomasi, M., Diaz, J., Ros, E., and Hulle, M. V. (2012). A comparison of FPGA and GPU for real-time phase-based optical flow, stereo, and local image features. *IEEE Trans. on Computers*, 61:999–1012.

Read, J. and Cumming, B. (2007). Sensors for impossible stimuli may solve the stereo correspondence problem. *Nature neuroscience*, 10(10):1322–1328.

Rodrigues, J. and du Buf, J. (2009). Multi-scale lines and edges in V1 and beyond: brightness, object categorization and recognition, and consciousness. *BioSystems*, 95:206–226.

Sanger, T. (1988). Stereo disparity computation using gabor filters. *Biol. Cybern.*, 59(6):405–418.

Solari, F., Sabatini, S., and Bisio, G. (2001). Fast technique for phase-based disparity estimation with no explicit calculation of phase. *Electronics Letters*, 37(23):1382 –1383.