# Improving the Vision of Magic Eyes: A Guide to Better Autostereograms

A.M. Bruckstein * , R. Onn ** , and T.J. Richardson *
* Bell Laboratories, Murray Hill, NJ 07974, USA and
** EE Department, Technion, I.I.T., Haifa, 32000, Israel

### Abstract

*An autostereogram is a single image that has the capability to convey depth information in the same manner as a stereo pair. Given a depth profile, the autostereogram is completely characterized by a two-dimensional basic pattern (a vertical strip.) Some autostereograms are more easily perceived than others depending on the basic pattern chosen to produce them. In this paper we discuss this dependence in terms of the spectrum of the basic pattern. We conclude that samples of 1/f noise yield excellent basic patterns, making it easy for the viewer to lock-into the desired depth profiles and to perceive depth in a stable way.*

## 1: Introduction - On Seeing Depth

The world around us is three dimensional, but eyes and cameras can only see planar projections of spatial scenes. Nevertheless, the third dimension can often be inferred from two dimensional images. Occlusion and prior information on object shapes provide depth information even in single images, and so do shadows and shading. Stereo vision provides a stronger effect enabling us to perceive depth: in stereo, depth is inferred locally from the slight differences in the images of the same scene produced by two horizontally displaced sensors (the eyes). Stereo vision is quantitative, in the sense that the binocular observer is able to evaluate the relative depth of almost all visible objects in a scene, a capability extensively exploited in geodesy. This quantitative depth perception depends on binocular disparity, and disappears when one of our eyes is closed; the visual system always fuses the available images so that a single image is perceived, with or without a sense of depth.

Depth evaluation in stereo vision is independent of the other visual cues that are usually present when viewing the world around us. This fact was shown by Julesz a long time ago, [Julesz,64], via a series of landmark experiments with random dot stereograms, i.e., pairs of similar images consisting of randomly placed dots in the plane, one of them having part of the dots displaced to encode depth. Depth is perceived when such image pairs are (simultaneously) presented to the two eyes of an observer. Independence of the stereo depth perception from other common depth cues was indeed to be expected from the observation that camouflaged objects, invisible in single images often become readily apparent in stereo pairs of images. It is the aim of camouflage to cover objects with a pattern that makes them appear to fuse with the background, their outlines or edges being completely obscured in monocular, or distant views. As in the case of camouflage, a random dot stereogram corresponds to a special (conceptual) coloring of the (imaginary) height profile/object surface with randomly spaced spots or dots, so that no information about the depth profile exists in a single image.

Let $x, y, z$ be coordinates for 3D space. To simplify the discussion we will assume that the depth of the surface to be viewed can be represented as a positive function $z = \varphi(x, y)$ and that no occlusions occur. This function will be referred to as the *depth profile*. From the depth profile we can easily construct a stereogram that allows the viewer to perceive it when his eyes are placed at, say $(-x_0, 0, -z_0)$ and $(x_1, 0, -z_0)$. We simply color the surface with some pattern $A(x, y)$ and project this pattern onto the plane $z = 0$ in two ways to produce two images $I_L(x, y)$ and $I_R(x, y)$ corresponding to the two different views of the surface as seen from $(-x_0, 0, -z_0)$ and $(x_1, 0, -z_0)$. In the sequel, we shall artificially assume that the surface is viewed by eyes located far way, with viewing directions of 90 and 45 degrees, i.e., $z_0 \to \infty$, $x_0 = 0$, $x_1/z_0 = 1$. This assumption lets us deal with parallel projection rather than perspective, simplifying the geometry considerably without sacrificing the essential features of the stereo imaging process. In fact, there is another depth profile $\varphi'$ which, when projected onto the two eyes, produces the same images as the extreme projection of $\varphi$ described above. It is this $\varphi'$ which will actually be perceived.

Stereo matching occurs in the horizontal direction, corresponding to the line through the two eyes. Therefore, images will be viewed as collections of rows, one corresponding to each $y$ -coordinate. For any particular $y_0$ we have to consider a pair of 1D functions of $x$ , $I_L(x, y_0)$ and $I_R(x, y_0)$. For notational convenience, we shall cease to write the $y$ coordinates for the bivariate functions involved.

Under these simplifying assumptions, we have (see Figure 1):

$$I_L(x) = A(x) \quad \text{and} \quad I_R(x + \varphi(x)) = A(x) \tag{1}$$

Note here that $I_R(x)$ is specified implicitly, via the depth function. If the function $x \to x + \varphi(x)$ is invertible, and this clearly happens if $|\frac{d}{dx}\varphi(x)| < 1$, then we can readily generate $I_R(x)$ from $I_L(x)$. (In practice, when the two projections are in directions much closer, say 90 and 80 degrees, the condition for invertibility is much milder, i.e. the depth function can have a much steeper slope).

As we have seen, producing the illusion of depth by showing to our eyes two slightly different images can be easily understood. This understanding led to many interesting practical applications, like 3D photography, visualization techniques in computer graphics, methods of photogrammetry, etc.

*But, the illusion of depth can also be induced by showing both eyes the very same image, provided this image is carefully designed!* This brilliant idea, that first occurred to C. W. Tyler in the seventies, is based on the freedom to choose a coloring of the surface that will yield two *identical* images as a stereo pair, [TylerChang,77], [Tyler,83] and [TylerClarke,90]. Today, countless books, postcards, and posters are dedicated to those "magical" images that, at a first glance look like an almost regular planar pattern of shapes and colors, but a "deeper" and more "defocused" look at them causes a second, and often amazing, three dimensional image to suddenly appear, see e.g. [MagicEye,93]. Tyler called such images *autostereograms*. In the sequel we shall try to explain how these images are are generated, and to analyze the way they are interpreted by the viewer, an analysis that will lead us to some ideas on how to design autostereograms that are more easily and stably interpreted as a three-dimensional image by the viewer. Although many papers in the literature addressed the topic of explaining and efficiently generating (mostly binary, or dot based) autostereograms, see e.g. [ThiIngWit,94], [TerTer,94], the issue of autostereogram

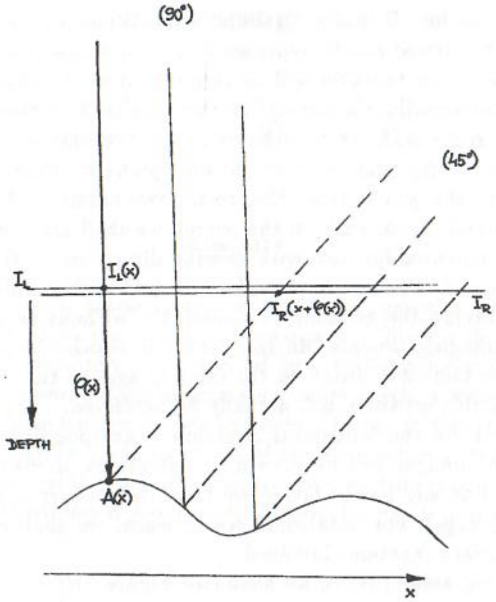

Figure 1. Stereo projection

design for easy interpretation seems to have never been raised and discussed.

## 2: Interpretation and Design of Autostereograms

As we have seen, an autostereogram is the image that would be obtained as two different projections of a suitably colored depth profile. From Equation (1) it becomes clear that the equality of $I_L(x)$ and $I_R(x)$ forces a coloring that obeys

$$A(x) = A(x + \varphi(x)) \text{ for all } x . \tag{2}$$

Referring to Figure 2, where again the projections are taken to be parallel at 90 and 45 degrees, and a single horizontal line of the image and depth profile are considered, we readily recover geometrically, that the image $I(x) = I_{R/L}(x)$ has to obey the basic functional equation inherited from $A(x)$ ,

$$I(x) = I(x + \varphi(x)) \text{ for all } x . \tag{3}$$

If the depth profile obeys the slope-limiting condition $|d\varphi(x)/dx| < 1$,[1] then it is easy to see that $I(x)$ is determined everywhere from its values on $x \in [t, t + \varphi(t))$ for any value of $t$. Indeed, all forward and backward iterations of the transformation $T : x \rightarrow x + \varphi(x)$, for any starting point $x_0$ have exactly one representative inside any interval of the type $[t, t + \varphi(t))$ . Thus, we are free to choose $I(x)$ over an arbitrary such interval and the rest of the stereogram (line) is then completely determined. We shall refer to $I(x)$ restricted to this (arbitrary) *basic interval*, as the *basic pattern* of the autostereogram. Note that the full autosterogram image will comprise a set of such parallel profiles (for various $y$'s)-representing horizontal lines in the image.

[1] It is possible to have perceivable discontinuities in depth in autostereograms but depth information is lost in the portion viewed by only one eye. This complicates the situation in a way which has little bearing on the subject of this paper.

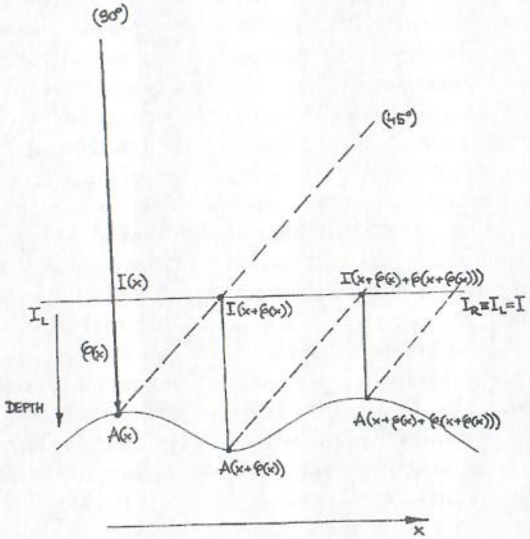

Figure 2. Auto-stereo projection

The freedom to choose the basic pattern can obviously be exploited to get a variety of visual effects and this freedom was indeed exploited, quite amazingly, in commercializing the autostereographic images. Interestingly however, depth is much more easily perceived in some autostereograms than in others. In fact there are autostereograms that satisfy the rules outlined above, and uniquely determine $\varphi$, and yet do not produce any depth perception at all. Hence, the following pair of questions arises naturally:

1. What is the underlying mechanism by which we perceive depth in such images?
2. How should we design basic patterns to make it easy for the viewer to perceive the third dimension?

As with almost any question about how biological mechanisms work, we cannot provide definite answers to the first question above. Unfortunately we can not answer Question 2. quantitatively without at least inferring a partial answer to Question 1. In Section 3 we develop a simple model for stereo vision, consistent with the available psychophysical evidence, which can be applied to autostereograms. Researchers studying human vision have proposed various models for stereo vision. The "squared differences" model we consider here was discussed by Sperling [Sperling,81] and Arndt et. al. [ArndtMallotBülthoff,95].

The combined use of this model with scale space arguments appears to be novel.

Recovery of the 3D (usually, depth) information from an autostereogram involves transcending the immediate, obvious and "planar" interpretation of the image seen. The Magic-Eye books, and the many similar books and posters now available, try to help viewers by guiding them to focus beyond the surface of the image (e.g.: "look at your own reflection in the window" on the other side of which an autostereogram image was posted, or "try to merge two feature points provided on the margins"). All of these are attempts to lead the human visual system toward a consistent, second peak of some locally defined matching process. To perceive the autostereogram, what we need to do is to correlate $I(x)$ with $I(x + \Delta(x))$ where $\Delta(x) \simeq \varphi(x)$, rather than being satisfied with the the perfect match obtained when correlating the image with itself at the same point in space, i.e. at $\Delta(x) = 0$, a match that fully supports the planar interpretation. Therefore, we can safely assume that a *local* correlation and matching process is at work, and that depth is inferred from the displacements at which matches were detected.

Let $I(x)$ be an autostereographic image. Let us construct a bivariate function (in fact trivariate, but remember that for the time being we work on a line-by-line basis!), $\Lambda(x, \tilde{x})$ that indicates how well $I(x)$ locally matches $I(\tilde{x})$.[2] This, clearly symmetric, bivariate function will have a high ridge along the diagonal corresponding to the flat image interpretation, since $I(\tilde{x})$ obviously matches $I(x)$ for $\tilde{x} = x$, but it should also have a very high ridge along the curve $\tilde{x} = x + \varphi(x)$. Similarly, we should have ridges along the curves $\tilde{x} = x + \varphi(x) + \varphi(x + \varphi(x))$, etc... The behavior of the 'surface' represented by $\Lambda(x, \tilde{x})$ clearly depends on how we define the local matching of images and on the particular choice of the autostereogram's basic pattern. Suppose that an oracle provided us with the physiologically correct matching function and the process used by the brain to find the best matching. Then, for a given depth profile, it would make sense to ask for the *optimal* autostereogram for a given depth profile. 'Optimal' here shall be interpreted in the sense that the basic pattern chosen yields sharp and high ridges along the curves $\tilde{x} = x$ and $\tilde{x} = x + \varphi(x)$, (and, inevitably, its iterates) and that the "domain of attraction" of the second interpretation should be as large as possible.

As a first exercise, suppose we are told that the depth interpretation is based on a pointwise grey-level match indicator function computed by the brain, "disparity" curves of the type $\tilde{x} = f(x)$ being evaluated by integrating (accumulating) $\Lambda(x, \tilde{x})$ along them. Clearly, with

$$\Lambda(x, \tilde{x}) = \chi\{I(x) = I(\tilde{x})\} = \begin{cases} 1 & if\ I(x) = I(\tilde{x}) \\ 0 & if\ I(x) \neq I(\tilde{x}) \end{cases} \tag{4}$$

the "correct" disparity curves carry a constant distributed weight along them ( $\Lambda(x, x) = \Lambda(x, x + \varphi(x)), ...etc$ ) but, depending on how $I(x)$ was designed, we might have other such ridges too. Furthermore, if the basic pattern of $I(x)$ is one-to-one then no additional ridges will exist, so, from the point of view of this matching function, all stereograms with one-to-one basic patterns will be equally good. However, a glance at the examples of Figure 3 clearly indicate that this matching function fails to capture some important features of the visual system's interpretation of autostereograms. Both basic patterns appearing in Figure 3 are one-to-one: The first is a ramp while the second is obtained by randomly permuting block-wise the ordinate of the ramp function. 3D interpretation turns out to be greatly facilitated by richness of detail and edginess in the basic pattern. One-to-one-ness,

[2] In reality the eye does not process horizontal lines independently: this dependence will be discussed and incorporated into our model in Section 3.

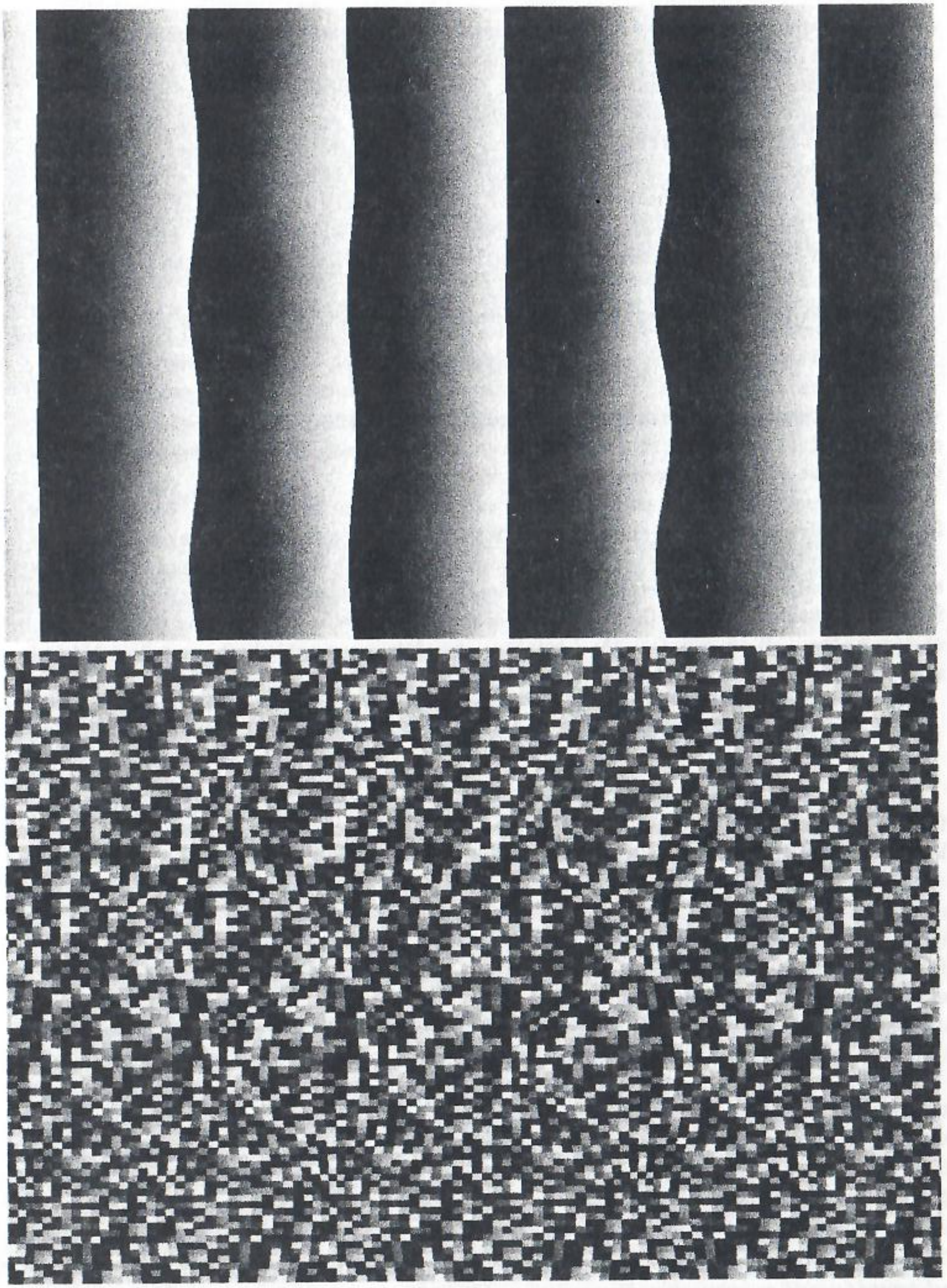

**Figure 3. Ramp function and random permutation of ramp function**

ensuring no spurious grey-level matches, is clearly less important than some measure of variability in the basic pattern. Stereo interpretation certainly involves not only local gray-level correspondences but also, more importantly, matches of regions with similar gray-level gradients, and similar average grey levels, matches of edges, of points and blobs, and, perhaps, even matches of complex 2D shapes, standing out as local features or tokens. Stereograms and autostereograms work over a wide range of input image types, from complex images requiring the matching of intricate, colored shapes, to very simple binary input images (and Tyler's first autostereograms were indeed black-and-white images.) These facts are also reflected in the many theories that have been put forward to explain the stereo vision process.

Let us next consider a class of slightly more complex matching functions[3] given by

$$\Lambda(x,\tilde{x}) \;=\; f([I(x) - I(\tilde{x})]^2) \qquad (5)$$

where $f$ is some smooth monotonically decreasing function satisfying $f(0) = 1$ and $\dot{f}(0) < 0$, e.g., $f(z) = 1/(1+\lambda z)$ or $f(z) = e^{-\lambda z}$. Such functions measure the local grey level distance too, but slight differences are better tolerated by them.

The "correct" disparity curves here also carry a uniform weight of one, but deviations from these curves yield more graceful degradations. The ridges corresponding to the curves $\tilde{x} = x$ , $\tilde{x} = x + \varphi(x)$ , etc... have downward slopes with steepness determined by the second derivative in the direction perpendicular to the curve. But along the curve the function is constant ( $= 1$ ), therefore sharpness of the ridges is expressed by the Laplacian of $\Lambda(x,\tilde{x})$ there. If we calculate the Laplacian at points on curves where exact matching of grey-levels occurs, e.g. on the curve $\tilde{x} = x + \varphi(x)$, we obtain

$$\nabla^2\Lambda(x,\tilde{x}) \;=\; 2\dot{f}(0)\left\{[\frac{d}{dx}I(x)]^2 + [\frac{d}{d\tilde{x}}I(\tilde{x})]^2\right\} \qquad (6)$$

Hence, the squared first derivatives of the basic pattern control the shape of the matching function along and around the disparity curves, and in fact at all places where the grey levels match along some curve. So we should (under the assumption that the $\Lambda(x,\tilde{x})$ under consideration is the correct one, and that our goal is to provide the sharpest ridge possible) design basic patterns with high derivatives almost everywhere and as few accidental matches (that do not correspond to desired disparity curves) as possible. Large first derivatives result from edges, hence, a basic pattern with many edges ensures large first derivatives almost everywhere and sharp maxima along the desired disparity curve. This seems to explain part of the results presented earlier (in Figure 3), although we do not yet have an understanding of the behavior of this matching function in between the ridges, for general basic patterns.

This class of matching functions indicates that local maxima are quite sharp if the patterns are rich, but, since the images have finite dynamic range, there will necessarily be many spurious matching points (and perhaps even curves) that will be equally sharp. Therefore, it would be advisable to have several basic patterns encoding each line of the depth pattern so that all of them will have consistent and sharp maxima along the correct disparity curves, but the spurious peaks located at *random* places in the areas in-between.

Then, in the actual matching process these spurious peaks would be averaged out. This calls for the use of samples of a noise process for the basic pattern, either in a perhaps far-fetched idea of using time-varying autostereograms (that would have to rely on temporal averaging in the visual system), or in static images, by exploiting the readily-available second dimension in the image plane (the $y$ -direction that we so conveniently disregarded until now!). Indeed, we can safely assume that the depth profile does not vary too fast in the $y$ -dimension, and use several consecutive lines of the autostereogram to encode the same (or slowly changing) depth pattern with a series of samples of a random-process, used to generate the basic patterns. (Physiologically, this amounts to assuming that stereo depth involves a coarsening of resolution.) In this context the question that remains to be answered is: what type of noise processes have the potential to yield good visual results?

Here again, we shall have to postulate the type of matching process that is performed by the visual system. We shall assume that the matching function is, in this case, dependent on averages of squared gray-level differences, over the samples of the process. Therefore, we define

$$V(x,\tilde{x}) \;:=\; E_\omega(I(x) - I(\tilde{x}))^2 \;=\; \mathbf{R}(x,x) + \mathbf{R}(\tilde{x},\tilde{x}) - 2\mathbf{R}(x,\tilde{x}) \qquad (7)$$

where $\mathbf{R}(x,\tilde{x})$ is the autocorrelation of the process $I_\omega(x)$ , whose samples are the lines of the autostereogram. If the process $I_\omega(x)$ is defined as the extension of a portion of a stationary process sampled over a basic interval, say, $[0,\varphi(0))$ , we obtain

$$V(x,\tilde{x}) \;=\; 2\mathbf{R}(0) - 2\mathbf{R}(B(x) - B(\tilde{x})) \qquad (8)$$

where $B(x)$ and $B(\tilde{x})$ are the $x,\tilde{x}$ "back-projected" into $[0,\varphi(0))$ according to the way $I_\omega(x)$ was extended beyond $\varphi(0)$. If we define the matching function to be

$$\Lambda(x,\tilde{x}) \;=\; f(V(x,\tilde{x})) \qquad (9)$$

we see that it will have ridges at the correct disparity curves and its behavior around these ridges will be determined by the autocorrelation function of the process used to build $I_\omega(x)$. In particular, if we make sure that the autocorrelation will only attain the maximal value of $\mathbf{R}(0)$ at zero and decay very steeply afterwards, we can shape the matching function into an approximate indicator function. In fact the Laplacian of $\Lambda(x,\tilde{x})$ at the ridges of local maxima, where $B(x) = B(\tilde{x})$ , is given by:

$$\nabla^2\Lambda(x,\tilde{x}) \;=\; 2\,\dot{f}(0)\,\mathbf{R}''(0)\left\{[\frac{d}{dx}B(x)]^2 + [\frac{d}{d\tilde{x}}B(\tilde{x})]^2\right\} \qquad (10)$$

showing that the shape of the ridge is controlled by the shape of the signal autocorrelation around the origin (we clearly do not have control over the depth-profile-dependent back-projection operator $B$ !). This result is derived under the assumption that $\mathbf{R}(\tau)$ is smooth about the origin and, being symmetric, it has zero first derivative there.

This discussion seems to indicate that we should favor processes with large values of $\mathbf{R}''(0)$. For example, the choice of a completely uncorrelated, white noise for the process generating the basic pattern will lead to very sharp local maxima at the correct disparities, meaning good behavior for autostereogram interpretations, provided we can keep the visual system "locked" into the various possible 3D interpretations. But locking into any particular depth profile, except the trivial one, can be, in this case, extremely hard. The matching function will not direct (via a hill climbing process) the visual interpretation toward any

[3] In [Sperling,81] and [ArndtMallotBülthoff,95] the same model is considered. Since human stereo does not depend strongly on correct normalization, the model is probably not, strictly speaking, correct. A more likely model is that of correlation. However, assuming correct normalization, minimizing squares differences approximates well maximizing correlation and the former affords a better mathematical framework because of the uniformity of the matching function along the correct match.

of the secondary maxima. An example of an autostereogram produced with white noise is given in Figure 4. We note that most of the random dot stereograms that were presented in the literature have the appearance of spatial white-noise and were indeed generated using random number generators and thresholds in quite a straightforward manner.

In the next section we present a hypothetical model for the matching process that is based on the common belief that the visual system processes images simultaneously at several scales. The images that are presented to us are "filtered" by several low, or band-pass filters that can be assumed to yield a "pyramid" of coarser and coarser, i.e. more and more blurred, images. It can be argued therefore that, in order for an autostereogram to look good and be easily interpretable, we need to have:

1. images that will look as homogeneous as possible over the entire span of spatial coordinates, in spite of the special way they were generated (which means that the random processes generating the autostereogram lines should be scaling invariant), and
2. images that lead to strong peaks of the image autocorrelation located at the correct disparities, but with with not too narrow "basins of attraction", about them at the coarse scales. In fact, it would help having basins of attraction tuned to the spatial scale, that become narrower as we go from the coarser to the finer scales, in order to direct the interpretation process (via hill-climbing on the corresponding matching functions!) toward the maxima located at the correct disparity curves.

To ensure the appearance of spatial homogeneity (as well as scale-space homogeneity) we would like to have a scale-invariant stochastic process generate the basic pattern, since the depth function locally scales the basic strips (in fact nonlinearly!) to produce the entire image. White noise would again be a reasonable candidate for this, but the requirement of having autocorrelation peaks with basins of attraction widening at a reasonable rate with the scale parameter is not met by a process with $\delta(\tau)$ autocorrelation.

Let $\mathbf{R}_\sigma(\tau)$ be the autocorrelations of the processes obtained when the basic pattern process is low-pass filtered to effective width $\sigma$ . We can analyze the behavior of $\mathbf{R}''_\sigma(0)$ as a function of $\sigma$ for various types of processes. It is seen that a white noise basic pattern leads, with decreasing $\sigma$, to a very quick narrowing of the peaks of the corresponding scale space of matching functions, $\Lambda_\sigma(x,\tilde{x})$ , while a noise whose spectrum that decays like $1/f^2$ in the frequency domain provides too slow a sharpening of the peaks with a decreasing scale parameter. A noise process that has $1/f$ behavior over the frequency range relevant to visual perception seems to be ideally suited for our needs. Indeed, $1/f$ -type noise has the property of selfsimilarity under scalings, needed for spatial homogeneity, and long-range correlation tails that will correctly guide the process of locking into the various depth interpretations!

To substantiate the above claim in a simple case, consider a constant depth profile. Then, $I_\omega(x)$ is a periodic process.[4] Hence the samples of the process can be described by a Fourier series as follows

$$I_\omega(x) = \sum_{i=0}^{\infty} a_i \cos(iw_0x + \phi_i)$$

where $\phi_i$ are i.i.d. random phases distributed uniformly over $[0, 2\pi)$, and $a_i$ are positive random variables too. The (periodic) auto-correlation of this stationary process is given by

[4] The constant depth case is rather trivial and is known in the stereo vision literature as the "wall-paper" effect for periodic patterns (see [Ittelson,60].)

$$\mathbf{R}(\tau) = \frac{1}{2}\sum_{i=0}^{\infty} E(a_i^2)\cos(iw_0\tau)$$

Now assume that we have a scale-space of filtered versions of the process $I_\omega(x)$ so that $I_\omega^\sigma$ is obtained by cutting off frequency components beyond $w_0/\sigma$. Then we have

$$\mathbf{R}_\sigma(\tau) = \frac{1}{2}\sum_{i=0}^{\sigma^{-1}} E(a_i^2)\cos(iw_0\tau)$$

and therefore,

$$\frac{\partial^2}{\partial(\tau/\sigma)^2}\mathbf{R}_\sigma(\tau) = -\frac{1}{2}\sum_{i=0}^{\sigma^{-1}} E(a_i^2)(i\sigma w_0)^2\cos(iw_0\tau)\,.$$

Note that here we have normalized $\tau$ by $\sigma$ since $\sigma$ is the appropriate unit of length on scale $\sigma$; we should expand $\mathbf{R}_\sigma(\tau)$ in terms of $\tau/\sigma$. Let us consider a sequence of matching functions $\Lambda_\sigma(x,\tilde{x})$ corresponding to the filtered versions of $I_\omega(x)$. Since the peaks of the matching function $\Lambda_\sigma(x,\tilde{x})$ is controlled by $\mathbf{R}''_\sigma(0)$, we see that the ($\sigma$ normalized) peaks of $\Lambda_\sigma(x,\tilde{x})$ get narrower with decreasing $\sigma$ at a rate described by:

$$F(\sigma) := \sigma^2 R''_\sigma(0) = -\frac{1}{2}\sum_{i=1}^{\sigma^{-1}} E(a_i^2)(i\sigma\omega_0)^2\,.$$

If we choose $E(a_i^2) \propto (i\omega_0)^{-\beta}$ (for $i \geq 1$) and let $\sigma \to 0$ we have

$F(\sigma) \to \sigma^{-1}$, for $\beta = 0$ ("white" noise),
$F(\sigma) \to$ constant, for $\beta = 1$ ("$1/f$" noise),
$F(\sigma) \to \sigma$, for $\beta = 2$ ("$1/f^2$" noise),
$F(\sigma) \to \sigma^2 \ln(1/\sigma)$, for $\beta = 3$ ("$1/f^3$" noise),
$F(\sigma) \to \sigma^2$, for $\beta > 3$ ("$1/f^{3+}$" noise).

Hence we see that with $1/f$ noise the peaks (normalized to length scale $\sigma$) retain essentially constant normalized width in scale space. More generally, we will have $\mathbf{R}_\sigma(\tau) \simeq \mathbf{R}(\tau/\sigma)$. This is desirable property for many reasons. We hypothesize that, whatever the matching mechanism is, it is invariant across scale. When $\tau$ is on the order of $\sigma$ we expect $\mathbf{R}_\sigma(\tau)$ to be the operative correlation. With $1/f$ noise we see that the width of the peaks of the matching function are directly proportional to scale. Suppose $\tau$ is adapted according to some hill climbing process in scale space. Then, in this case, for $\tau \ll \sigma$ we are well within the peak of matching function on scale $\sigma$ so that scale will contribute little to the correction in $\tau$. Having $\sigma \simeq \tau$ places $\tau$ on the steep portion of the peak, hence a strong indication of the appropriate correction to $\tau$ is generated. For $\tau \gg \sigma$ we expect to be outside the domain of attraction and small, perhaps random, corrections to $\tau$ are indicated. Thus $1/f$ noise appears ideal from the point of view of obtaining stable convergence to the peak over the widest possible range. For $\beta > 1$ we expect convergence to break down on small scales, hence resolution will be lost. For $\beta < 1$ we expect convergence to break down on large scales, so the domain of attraction of matching will be reduced and the autostereogram will be harder to perceive.

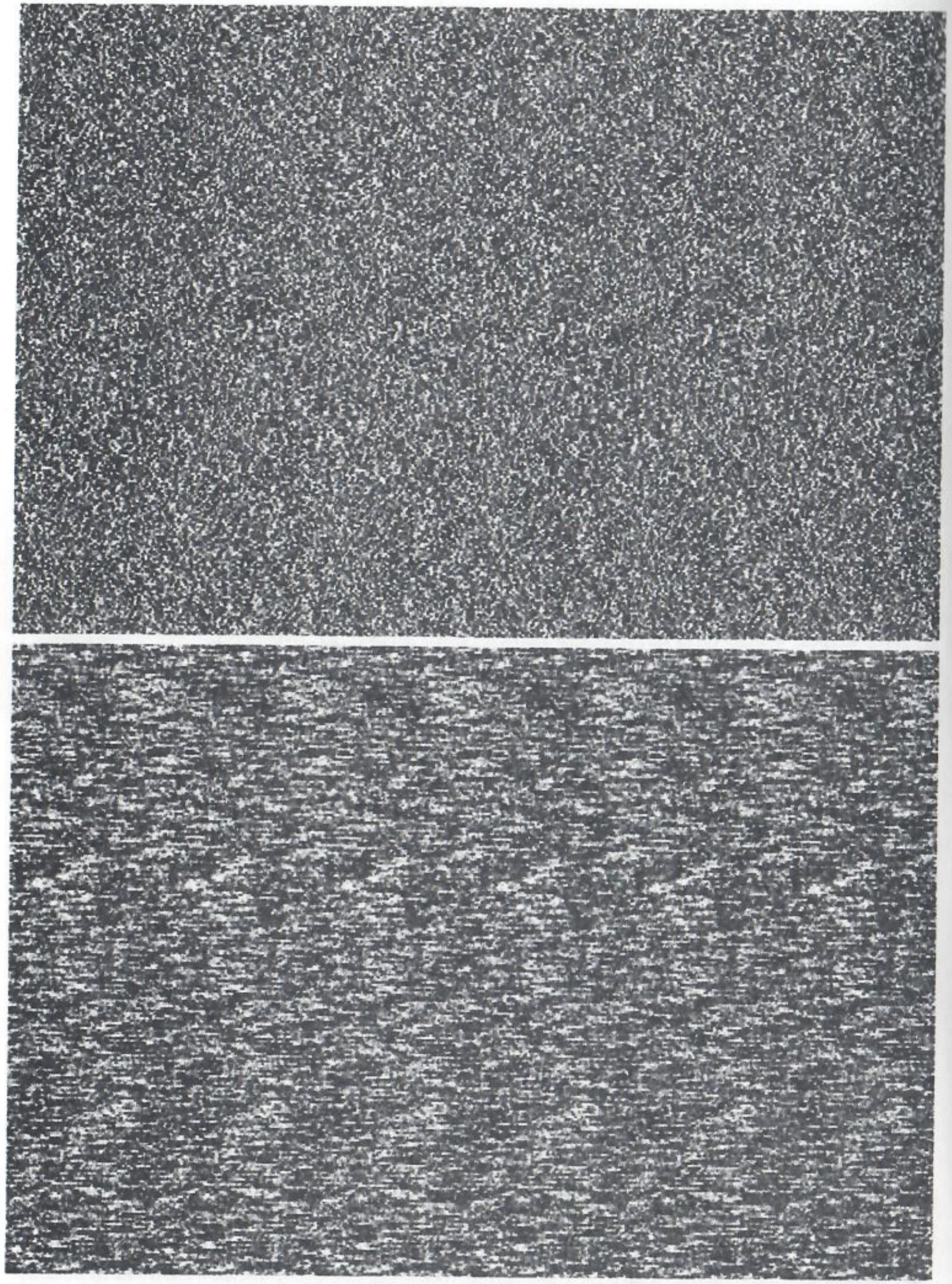

Figure 4. Independent lines of $1/f^0$ (white) noise and $1/f$ noise

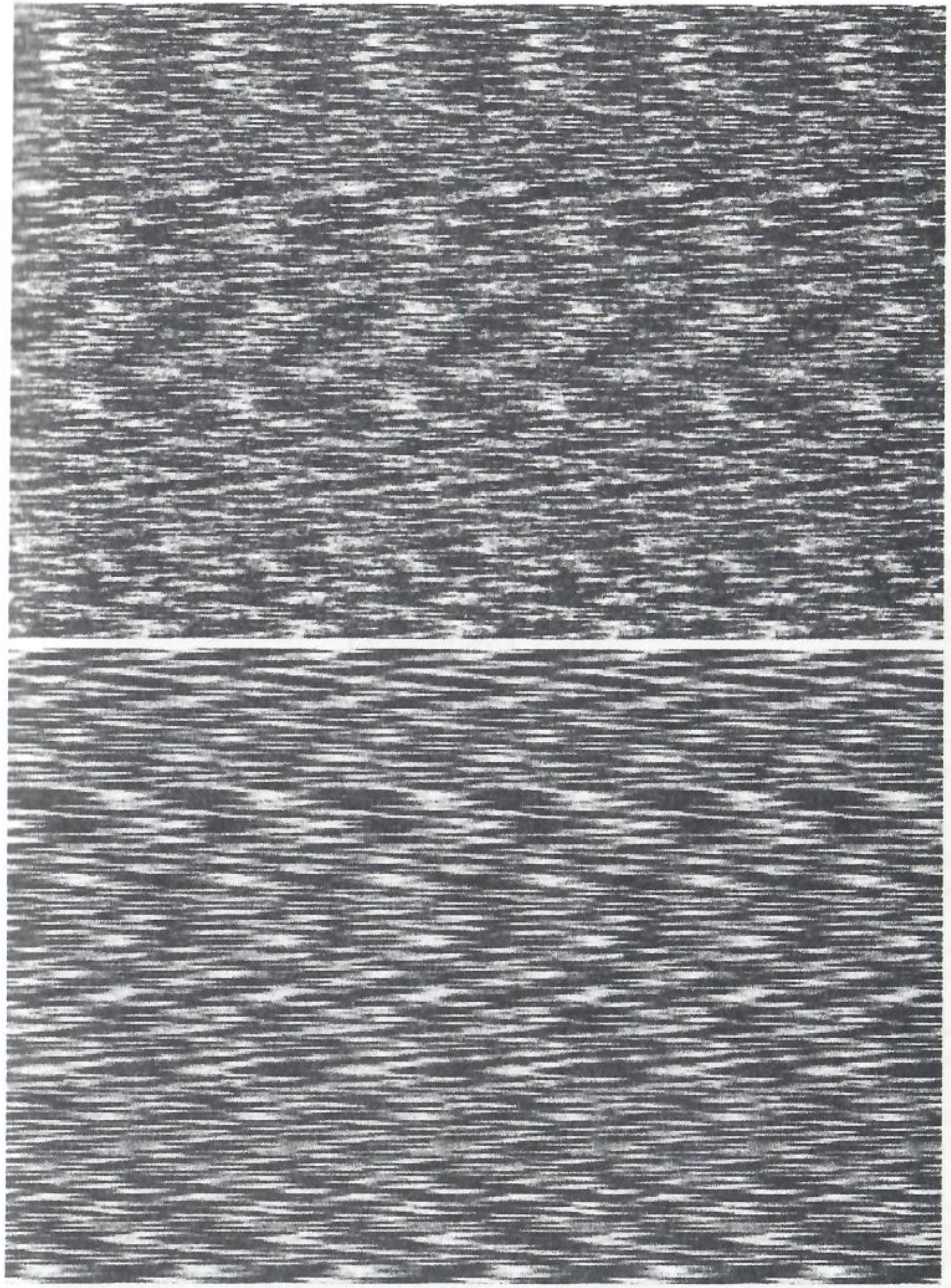

Figure 5. Independent lines of $1/f^2$ (white noise) and $1/f^3$ noise

Figures 4 and 5 present a series of autostereograms with independent line process having power spectra $1/f^n$ for $n = 0, 1, 2, 3$. Note the degradation in resolution of depth for $n = 3$. It is difficult to choose between $n = 1$ and $n = 2$ for overall quality. In the next Section we will further elaborate on the role of scale-space in stereo vision and in perceiving autostereograms in particular. There we will conclude that a 2D $1/f$ process is desirable; independent line processes of $1/f^2$ noise actually approximate this better than independent line processes of $1/f$ noise.

## 3: From Scale Space to Autocorrelation

So far we made the assumption that the visual system has a way to compute the ensemble averages of the horizontal line processes, the averages being needed to effectively evaluate the matching functions $\Lambda(x, \tilde{x})$ that lead to 3D interpretations. It was suggested that use could be made of the vertical direction to mimic ensemble averaging: By having different samples on different lines vertical averaging of line by line products could mimic ensemble averaging. The reader might object to this idea on the basis that not enough independent lines will enter the depth calculation to simulate ensemble averaging. It turns out that a simple frequency decomposition of the image, an operation which is widely believed to be performed by the visual system, can facilitate this process. This can be seen even in 1D. Consider, as before, an image line

$$I(x) = \sum_i a_i \sin(iw_0 x + \phi_i)\,.$$

Let us assume that we have independent lines of this form where the $\phi_i$ are independent and uniform and, for simplicity, the $a_i$ are deterministic (i.e., the same from line to line). Let $I_i(x)$ denote $a_i \sin(iw_0 x + \phi_i)$. The calculation of the correlation takes the form

$$E(I(x)I(x+\tau)) = E(\sum_i I_i(x)I_i(x+\tau)) + E(\sum_{i \neq j} I_i(x)I_j(x+\tau)\,,$$

where the expectation is taken over the random phases of the sinusoids. The expectation eliminates the second term. However, if we wish to approximate the expectation by a sum over independent samples then, indeed, the number of samples required can be quite large. This is due to the relatively large number of essentially independent terms which need to be averaged out. However, if each eye-sensor were able to first extract $I_i$, by Fourier analysis, then the brain could compute $\sum_i I_i(x)I_i(x+\tau)$ directly thereby significantly reducing the number of independent samples required to approximate the ensemble average. Furthermore, since $I_i(x)I_i(x+\tau) = a_i^2 \cos(iw_0\tau) + a_i^2 \cos(iw_0(2x+\tau) + 2\phi_i)$, the second, spatially (i.e. $x$) dependent, term could also easily be removed by low pass filtering with a cutoff at or above $iw_0$. Note that a lower cut-off is not feasible since this would result in extreme loss of resolution in depth. (Recall that in reality all calculations are local.) Fourier analysis as indicated here is not a realistic assumption for the eyes but the point is this: band-pass decomposition of the images, correlation within bands, and low-pass filtering of the products prior to recombining can significantly reduce the burden of ensemble, i.e. vertical, averaging.

Let us elaborate a bit further on our ideal model. Suppose, as above, that the brain computes $I_i(x)I_i(x+\tau)$. It is reasonable to assume that averaging occurs before recombining over $i$. In this way we may conclude that the brain computes

$$\mathbf{P}^i(\tau) := E(I_i(x)I_i(x+\tau)) = \frac{1}{2}a_i^2 \cos(iw_0\tau)$$

separately for each frequency band, i.e. (ideally) for each $i$. Now, let us postulate that, in order to adjust $\tau$, $\mathbf{P}^i(\tau)$ is differentiated as a function of $\tau$. Since the length scale associated with frequency $w_0 i$ should be $1/(w_0 i)$ we define

$$D^i(\tau) := \frac{1}{iw_0}\frac{\partial}{\partial\tau}\mathbf{P}^i(\tau) = -a_i^2 \sin(w_0 i\tau)\,.$$

Note that this scaling is possible only if the averaging is done before recombining the various frequency bands, although the differentiation may occur after. Now, we will assume that the temporal derivative of $\tau$, $\frac{\partial\tau}{\partial t}$ is computed by the brain by summing the length normalized contribution from different frequency bands. We will therefore define,

$$D(\tau) := \sum_i D^i(\tau) = -\sum_i a_i^2 \sin(iw_0\tau)\,.$$

The question now is: what should we choose for $a_i$ in order to obtain good behavior for the $\tau$-adjustment rate. For simplicity, and to enable scale invariance, let us consider the possibility $a_i^2 \propto i^\beta$ for $i \geq 1$ and ask what to choose for $\beta$. Let $\omega_{\max}$ be the highest frequency actually used, then we have,

$$D(\tau) \simeq -(\text{const }) \int_{w_0}^{\omega_{\max}} \omega^\beta \sin(\omega\tau)\, d\omega = -(\text{const })\tau^{-1-\beta} \int_{w_0\tau}^{\omega_{\max}\tau} u^\beta \sin(u)\, du$$

In the case of interest, $w_0\tau \ll 1$ and $w_{\max}\tau \gg 1$, and $-2 < \beta < 0$ we have $D(\tau) \simeq$ (const) $\tau^{-1-\beta}$. Given a global constraint on the energy in our images, arising, for example, from limitations in the rendering, $\beta$ controls a trade-off between high resolution and the domain of convergence. This trade-off is balanced when we choose $\beta = -1$. Smaller values of $\beta$, i.e., $\beta < -1$ favor larger scales. Convergence slows as $\tau$ becomes small and resolution is reduced (because of noise.) For larger $\beta$, i.e., $\beta > -1$ we favor smaller scales, achieving high resolution but reducing the domain of convergence.

Once again, by invoking a scale invariance assumption we have concluded that a $1/f$ power spectrum is ideal for our basic patterns. But what properties should the 2D basic strip have ? We could, for example, let each horizontal line of the basic strip be an independent sample of a $1/f$ type noise process. The results of this idea are seen in Figure 4; if the eye processed each line independently and then, in the final "correlation" stage, averaged vertically over several lines, then, according to our analysis, this type of basic strip would be ideal. However, the eyes do not work in this, line by line, way. This is clearly indicated by the fact that a basic strip with 2D $1/f$ noise is significantly superior to line independent $1/f$ noise, as we shall see in the experiments.

A physiologically more plausible model would allow for the 2D filtering performed by the eyes prior to correlation. It is widely held that the visual system decomposes the images projected onto the eyes into various frequency bands. A standard model is isotropic band-pass filtering, i.e., the gain of the 2D frequency $w$ depends only on $|w|$. For stereo processing there might well be some anisotropy but not the extreme version considered earlier. For simplicity we will study the effect of 2D filtering by considering the isotropic model.

Let us now try to understand our previous models from the 2D point of view. To begin with we shall return to the model of 1D correlation of the images with vertical averaging of the product. We will later adjust the model to add other elements. For notational convenience we will let $I_\tau$ be defined by $I_\tau(x,y) := I(x+\tau, y)$. Let $\mathcal{F}$ denote the Fourier transform, i.e.,

$$\mathcal{F}(I)(w) = \mathcal{F}(I)(w_1, w_2) = \int\int e^{i(w_1 x + w_2 y)} I(x,y)\, dx\, dy\,.$$

Let $G(w)\,(= \delta(w_2))$ denote the transfer function of (ideal) vertical averaging. Our simple model reduces to computing the following,

$$\mathcal{F}^{-1}(G \cdot \mathcal{F}(II_\tau)) = \int\int (e^{i\langle u,\vec{x}\rangle}\mathcal{F}(I)(u))^* G(w-u) e^{iw_1\tau}(e^{i\langle w,\vec{x}\rangle}\mathcal{F}(I)(w))\, dw\, du \qquad (11)$$

where $\vec{x} = (x,y)$. We can introduce band pass filtering into this model as follows. Let $H^\sigma(w)$ denote the transfer function of some band-pass filter and assume $\sum_\sigma H^\sigma(w) = 1$. Then for each $\sigma$ the eyes/brain may compute

$$\int\int (e^{i\langle u,\vec{x}\rangle} H^\sigma(u)\mathcal{F}(I)(u))^* G^\sigma(w-u) e^{iw_1\tau}(e^{i\langle w,\vec{x}\rangle} H^\sigma(w)\mathcal{F}(I)(w))\, dw\, du\,. \qquad (12)$$

Here we have admitted the possibility that the filter $G$ may depend on $\sigma$. In the case that $H^\sigma \simeq \delta(|w_1| - 2\pi/\sigma)$, i.e. ideal horizontal band-pass filtering, and $G^\sigma$ introduces some low-pass filtering in the horizontal direction, then we reproduce the scenario described at the beginning of this section. Our 2D model would have $H^\sigma$ represent 2D bandpass filtering in a band near $2\pi/\sigma$, i.e. $H^\sigma$ passes frequencies $w = (w_1, w_2)$ with $|w| \simeq 2\pi/\sigma$. For simplicity we will assume $H^\sigma$ is ideal, i.e.,

$$H^\sigma(w) = \begin{cases} 1 & w \in \Omega_\sigma \\ 0 & w \notin \Omega_\sigma \end{cases}$$

where $\Omega_\sigma$ is some annulus $\{|w| \simeq 2\pi/\sigma\}$. The filter $G^\sigma$ should represent vertical averaging and also, perhaps, low pass horizontal filtering with cutoff near $|w| \simeq 2\pi/\sigma$.

We will study the effect of the 2D filtering via an illustrative example. We consider an image of the form

$$I(x,y) = \sum_i a_i \sin(iw_0 x + \phi_{ik})$$

where the $a_i$ are independent of $y$, $k$ is determined by $y \in [k\epsilon/w_0, (k+1)\epsilon/w_0)$, and $\phi_{ik}$ are uniformly random in $[0, 2\pi]$ and, for now, independent for each $i$ and $k$. This models an image in which pixels are vertically separated by $\epsilon/w_0$ and each horizontal line is independent.

If we examine the 2D spectrum of such an image we find that, roughly speaking, energy $a_i^2$ is distributed uniformly in a strip around the line segment $\{w_1 = iw_0, w_2 \in (-w_0/\epsilon, w_0/\epsilon)\}$ and its reflection $\{w_1 = -iw_0, w_2 \in (-w_0/\epsilon, w_0/\epsilon)\}$. Thus, the energy associated with horizontal frequencies gets smeared, approximately uniformly, across a wide range of vertical frequencies.

For the image above, when $G_\sigma$ is ideal vertical averaging we obtain

$$\mathcal{F}^{-1}(G_\sigma \cdot \mathcal{F}(I^\sigma I^\sigma_\tau)) = \frac{1}{2}\int_{\Omega_\sigma} |\mathcal{F}(I)(w)|^2 \cos(w_1\tau) dw \qquad (13)$$

where $I^\sigma$ denotes $\mathcal{F}^{-1}(H^\sigma \cdot \mathcal{F}(I))$. As before, we hypothesize that the eyes compute

$$D_\tau := \sigma\frac{\partial}{\partial\tau}\mathcal{F}^{-1}(G_\sigma \cdot \mathcal{F}(I^\sigma I^\sigma_\tau)) = -\frac{1}{2}\int_{\Omega_\sigma} |\mathcal{F}(I)(w)|^2 (\sigma w_1) \sin(w_1\tau) dw$$

This reveals two weaknesses associated with independent lines. If the eyes do 2D band pass filtering as suggested above, then low (absolute) frequencies will not have enough energy in them and higher frequencies too much. This explains, in part, why 2D $1/f$ noise is superior as a basic strip to independent lines of $1/f$ noise. The second undesirable effect of having independent lines is that putting energy in frequencies $(w_1, w_2)$ where $|w_2| \gg |w_1|$ is wasteful in the sense that stereo depth information is not carried by vertical components. Since the derivative above associated to $w = (w_1, w_2) \in \Omega_\sigma$ is, according to our model, scaled by $\sigma \simeq 1/|w|$ it is better, in the sense that the derivative will be larger, to have $w_1 \gg w_2$. The limiting case, $w_2 = 0$, is however undesirable since this requires phases $\phi_{ik}$ which do not vary over $k$, eliminating the tremendous value of vertical averaging. Thus we observe that there is a trade off between providing for the local averaging and maximizing horizontal signal.

A compromise is the following. We should let $\phi_{i,k}$ be positively vertically correlated, i.e. correlated in $k$, in a way which depends on $i$. We would like the randomness in the $\phi$'s to result in a 2D spectrum which is largely supported, roughly speaking, in the cone $|w_2| \leq |w_1|$. This can be achieved, for example, by letting $\phi_{ik}$ be a sample path from a random walk as follows,

$$\phi_{i,k+1} = \phi_{ik} + \gamma\xi_{i,k+1}/\sqrt{i}$$

where $\xi_{ik}$ are i.i.d. uniform on $[-1,1]$ (say) and where $\gamma$ is an appropriate constant independent of $i$ and $k$. In this way the 2D spectrum is also essentially $1/f$. The only remaining concern is whether vertical averaging can still effect cancellation of cross terms in the products $I^\sigma I^\sigma_\tau$. To see this we need to reconsider the derivation of equation (13). For simplicity let us assume that that horizontal frequencies are not spread across 2D bands. Then the phases of the cross terms take the form $\phi_{ik} \pm \phi_{jk}$. Thus, the cross term phases will behave like a random walk but the correlation length will be on the order of $1/i + 1/j$. Since, because of scale space filtering, we can assume $i \simeq j$ we see that the cross term phases fluctuate significantly more rapidly than the local length scale $\simeq 1/i$, and, therefore, vertical averaging would be able to effectively eliminate them.

Examples of images produced this way are given in Figures 6 and 7. Here we can see the role of vertical correlation. The figure at the bottom of Figure 6, a suitable vertically correlated $1/f$ noise pattern, is among the 'best' autostereograms we have been able to produce. Comparing Figure 6 with Figure 7 the superiority of $1/f$ to $1/f^2$ noise when the second dimension is properly taken into account is evident.

## 4: Discussion and Concluding Remarks

Autostereograms are natural generalizations of periodic patterns that were known to produce the so-called "wall paper phenomenon" documented in old books on visual perception. Ittelson, [Ittelson,60], describes this phenomenon as follows: "An observer stands a few feet distant from, and squarely facing, a wall covered with a regular, repeating pattern of small figures. By increasing the convergence of his eyes while observing the pattern on the wall, the observer will note that there are one or more amounts of convergence for

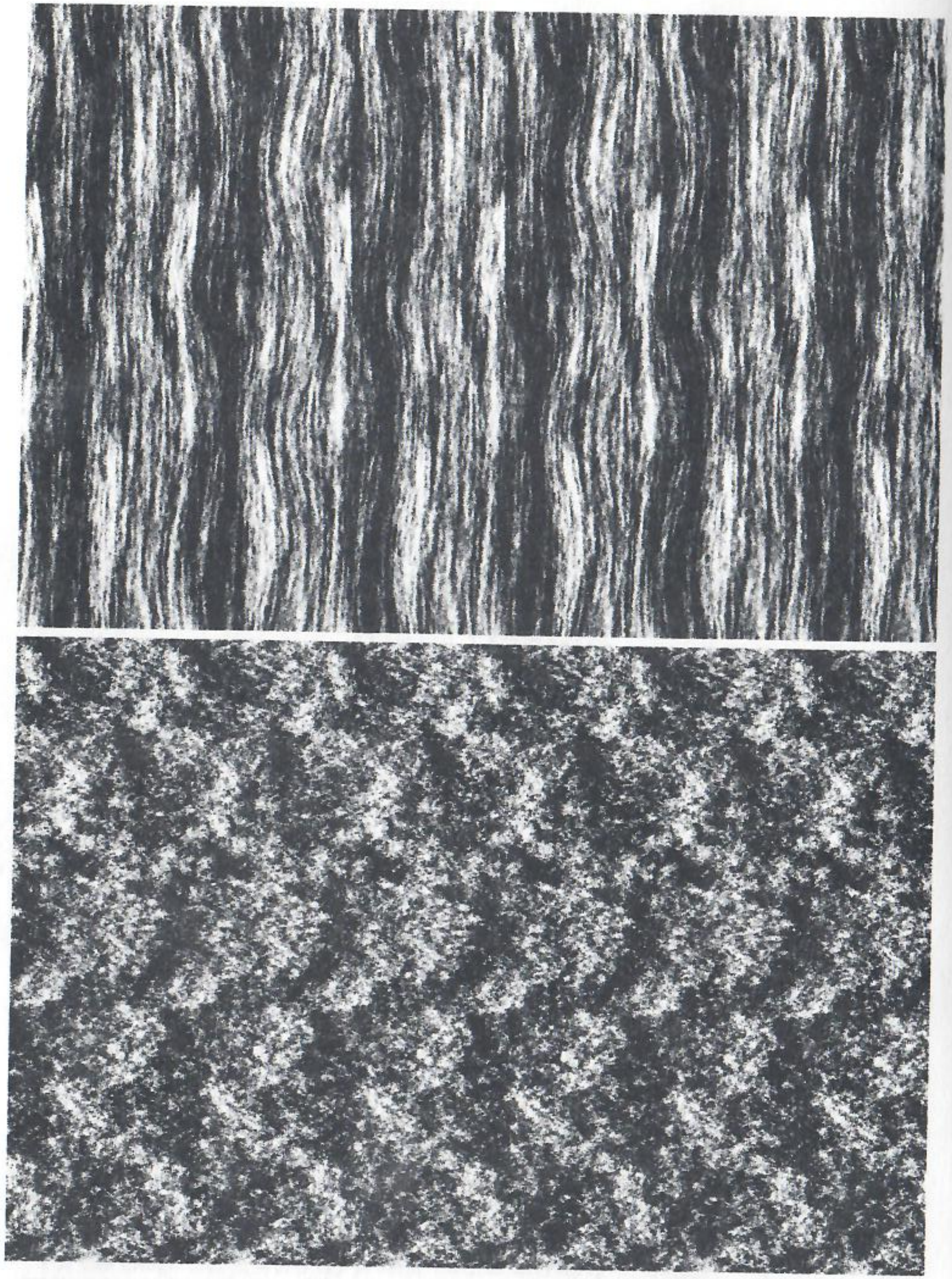

Figure 6. Patterns of $1/f$ noise with different vertical correlation

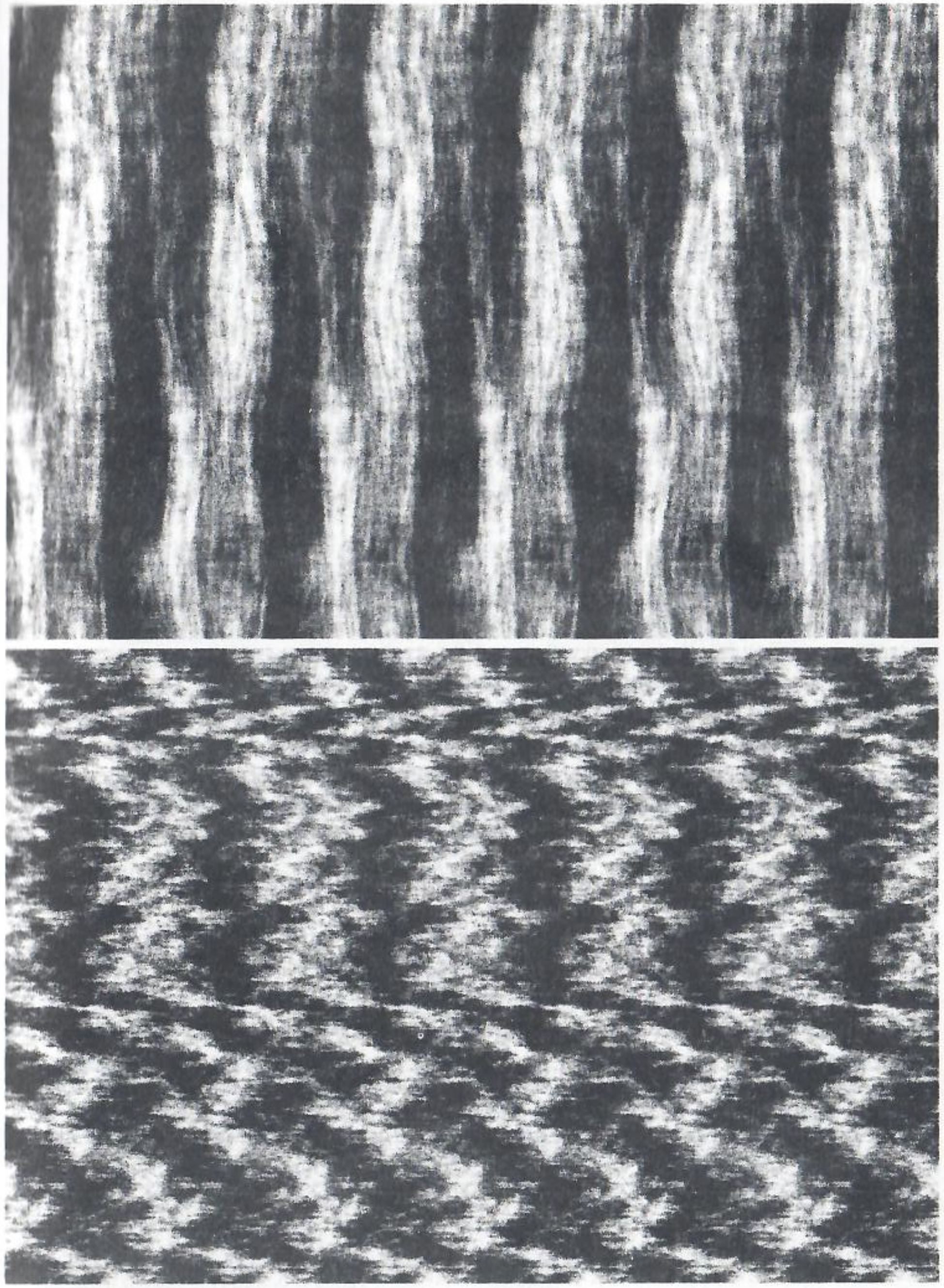

Figure 7. Patterns of $1/f^2$ noise with different vertical correlation

which fusion will be obtained, but with different, rather than the same, parts of the pattern fusing together. At the same time, the entire wall will appear to have moved nearer to the observer and become smaller. The same effect has been observed for a typewriter keyboard, postage stamps, and various other repeated figures."

In this short paper we explained the way autostereograms are produced for arbitrary depth profiles and posed the question of designing autostereograms for best visual interpretations. Then we analyzed possible simplistic mechanisms for depth recovery from the autostereographic images, and concluded with arguments that point toward $1/f$ -noise processes as excellent generators for basic autostereogram patterns. Further work on this topic is currently under way, analyzing the relationships between the processes chosen to generate basic patterns and the ease of locking into the 3D interpretations, under a variety of further physiologically motivated stereo interpretation models.

Much deeper analysis is required to solve the problem of optimal autostereogram designs for complex models of stereo perception, but we believe that $1/f$ noise will turn out to be universally good for these purposes. Easily perceived autostereograms might one day be viable alternatives for the effective display of three dimensional surfaces and data.

## References

[Julesz,64] B. Julesz, *Binocular Depth Perception without Familiarity Cues,* Science, Vol. 145, pp. 356-362, 1964.

[TylerChang,77] C. W. Tyler, J. J. Chang, *Visual echoes: the perception of repetition in random patterns,* Vision Research, Vol. 17, pp. 109-116, 1977.

[Tyler,83] C. W. Tyler, *Sensory Processing of Binocular Disparity,* in Vergence Eye Movements: Basic and Clinical Aspects, Butterworth, Boston, pp. 199-295, 1983.

[TylerClarke,90] C. W. Tyler, M. B. Clarke, *The Autostereogram,* Proceeding SPIE. Meeting on Stereoscopic Displays and Applications, Vol. SPIE 1256, pp. 182-197, 1990.

[ThiIngWit,94] H. W. Thimbleby, S. Inglis and I. H. Witten, *Displaying 3D Images: Algorithms for Single Image Random Dot Stereograms,* Computer, Vol. 27/10, pp. 768-774, 1994.

[TerTer,94] M. S. Terrel and R. E. Terrel, *Behind the Scenes of a Random Dot Stereogram,* American Math, Monthly, pp. 715-724, October 1994.

[Sperling,81] G. Sperling, *Mathematical Models of Binocular Vision,* SIAM-AMS Proc. Vol. 13, pp. 281-300, 1981.

[ArndtMallotBülthoff,95] P. A. Arndt,H. A. Mallot and H. H. Bülthoff, *Human stereovision without localized image features,* Biol. Cybern. 72, pp. 279-293, 1995.

[MagicEye,93] *N. E. Thing Enterprises, Magic Eye: A New Way of Looking at the World,* Michael Joseph Ltd, Penguin Group, 1993.

[Ittelson,60] W. H. Ittelson, *Visual Space Perception,* Springer Publishing Co, New York, pp. 123-127, 1960.

# Shape Recovery from Stationary Surface Contours by Controlled Observer Motion

Liangyin Yu Charles R. Dyer

Computer Sciences Department
University of Wisconsin
Madison, WI 53706

### Abstract

*The projected deformation of stationary contours and markings on object surfaces is analyzed in this paper. It is shown that given a marked point on a stationary contour, an active observer can move deterministically to the osculating plane for that point by observing and controlling the deformation of the projected contour. Reaching the osculating plane enables the observer to recover the object surface shape along the contour as well as the Frenet frame of the contour. Complete local surface recovery requires either two intersecting surface contours and the knowledge of one principle direction, or more than two intersecting contours. To reach the osculating plane, two strategies involving both pure translation and a combination of translation and rotation are analyzed. Once the Frenet frame for the marked point on the contour is recovered, the same information for all points on the contour can be recovered by staying on osculating planes while moving along the contour. It is also shown that occluding contours and stationary contours deform in a qualitatively different way and the problem of discriminating between these two types of contours can be resolved before the recovery of local surface shape.*

## 1: Introduction

Natural objects are full of textures of all kinds, providing qualitatively different cues about surface shape. Different kinds of texture require different methods for analysis. One kind of surface texture, stationary surface contours, constrains surface shape in a way very different from "blob-like"texture. These stationary contours are one-dimensional curvilinear markings on the object surface, which, unlike occluding contours, do not "slide" across the surface as the vantage point changes [4] and, hence, only constrain the surface along a single dimension like a strip for a smooth surface [10]. Consequently, stationary contours have been studied mostly in the context of qualitative surface characterization [5, 9, 12, 14]. In contrast, since the observation of Barrow and Tenebaum [1] that occluding contours constrain surface orientation uniquely even from a single viewpoint, this kind of contour has been the focus of considerable research to quantitatively characterize the surface from one-dimensional curvilinear features.

The support of the National Science Foundation under Grant No. IRI-9220782 is gratefully acknowledged.