How to See in the Dark: Photon Noise in Vision and Nuclear Medicine[a]

CHARLES S. PESKIN,[b] DANIEL TRANCHINA,[b,c] AND DIANA M. HULL[d]

[b]*Courant Institute of Mathematical Sciences*
New York University
New York, New York 10012

[c]*Laboratory of Biophysics*
Rockefeller University
New York, New York 10021

[d]*Medical Scientist Training Program*
Albert Einstein College of Medicine
Bronx, New York 10461

INTRODUCTION

Next time you meet a white cat in a dark hallway, take a close look at the amount of detail you can see. Depending on the amount of available light, the cat will appear in various degrees of blurriness. Seeing the texture of the fur is out of the question. Details such as ears are hard to make out. Sometimes the cat is just a ghostly blur and it is hard to tell which end is the head and which is the tail. Evidently, there is a relationship between the amount of light available and the spatial resolution that can be achieved in vision.

The experimental study by Rose[1] was the first to show clearly that the rod system in the human retina is often starved for photons. This implies that some limitations of rod-mediated visual performance are imposed by quantal fluctuations in the retinal image. A highly reflecting object, illuminated by a full moon on a clear night, will give rise to roughly 3 captured photons per rod per second in the retinal area with greatest rod density. Clearly, under these conditions, features in the retinal image which are on a scale similar to the spacing between rods will not be perceived reliably. This is particularly true if the viewed object is moving. For further discussion of experimental evidence concerning the significance of photon noise in vision, see the comprehensive review article by Shapely and Enroth-Cugell.[2]

To understand the limitation on visual resolution that is imposed by the photon noise consider the analogous situation in photography. If we look closely at a photographic image, we see that it consists of a number of grains. Each grain is the record of the capture of a photon. In fact, the information content of a photograph may be summarized by listing the positions of the grains. Now consider the problem of reconstructing an image from such a list. The brightness of the image should be the "density" of the grains, but the concept of density becomes elusive when we get down to

[a]This work was supported in part by The MacArthur Foundation, Grant MCS-8201599 from The National Science Foundation, and Grant 5 T32 GM7288-09 from The United States Public Health Service.

a length scale that is comparable to the typical distance between nearby grains. The obvious way to define density at a point is to surround the point by a patch of area, count the number of grains in the patch and divide by the area of the patch.

The question is how to choose the area of the patch. From the standpoint of spatial resolution, the patch should be as small as possible, since it is clearly impossible to resolve features in a smaller scale than that of the patch. From the standpoint of noise reduction, however, the patch should be as large as possible. The number of grains in the patch is a random variable (that is what we mean by "photon noise") given by the Poisson distribution. This distribution has the property that the standard deviation is the square root of the mean. To achieve an expected error of ±10%, for example, the number of grains in the patch has to be at least 100. At lower levels of light there are fewer photons per unit area, so the size of the region that is needed for a reliable estimate of the density is correspondingly larger. This explains the loss of resolution at low levels of light. The resolution has to be deliberately reduced because an attempt to maintain high resolution would result in an unacceptably noisy image.

We will show, however, that there is a better (albeit closely related) strategy for coping with the photon noise than merely pooling the data over regions of carefully chosen size. This strategy involves the design of an optimal linear filter for reconstructing an image from a record of photon absorption events. We believe that the retina is a closely related filter and that the theory of optimal filters has features that are strikingly reminiscent of the phenomenon of light adaptation in the retina. Moreover, we believe that such filters will turn out to be very useful in nuclear medicine, since nuclear medicine images are subject to considerable degradation by photon noise.

The mathematical theory of optimal filters that we use throughout this paper is not new. It was invented by N. Wiener[3] and has been widely used in the design of communication systems. Further references on this subject are those of Lee[4] and Morse and Feshbach.[5] The situation we consider in this paper does have the somewhat novel feature that the input to the filter is not the *sum* of a signal and a noise. Rather, the signal has been encoded in a manner that is intrinsically noisy: the observer measures the brightness of a scene only by counting the (random) photon events. The main contribution of this paper, though, is not in the theory but in the applications, which we believe are new.

OPTIMAL FILTER FOR THE PHOTON IMAGE

Let $u(\mathbf{x}) \geq 0$ be the expected number of grains (photons) per unit area at position $\mathbf{x}$ in a photograph of a scene. The actual grains are located at positions $\mathbf{X}_k$, where k is an arbitrary label (the order of the photons has no significance). Given $u(\mathbf{x})$, the $\{\mathbf{X}_k\}$ are generated by an inhomogeneous Poisson process with density $u(\mathbf{x})$.

There are many equivalent ways to define a Poisson process. The following two-step construction of the process is particularly convenient:

(i) Choose an integer N according to the Poisson distribution:

$$Pr\{N = n\} = \frac{\mu^n}{n!} e^{-\mu}, \quad (1)$$

where

$$\mu = \int u \, d\mathbf{x}. \quad (2)$$

(ii) Given N, choose $\mathbf{X}_k$, $k = 1, \ldots, N$, independently, from the probability density function

$$\rho(\mathbf{x}) = \mu^{-1}u(\mathbf{x}). \tag{3}$$

Remark: This definition assumes that $\int u d\mathbf{x} < \infty$. If not (for example if u = constant and the region in question is the entire plane), then break the region up into subregions where $\int u d\mathbf{x} < \infty$ and apply the foregoing definition in each subregion.

We have defined the stochastic process that generates the photon image $\{\mathbf{X}_k\}$ given the photon density function $u(\mathbf{x})$. Note that the quantity $u(\mathbf{x})$ is dependent on the illumination of the scene. A quantity that is more characteristic of the scene itself is the reflectance $r(\mathbf{x})$, $0 \le r(\mathbf{x}) \le 1$, which is the fraction of the incident light that is reflected at the point $\mathbf{x}$. Let I_0 be the expected number of photons per unit area of photograph when $r(\mathbf{x}) \equiv 1$; thus I_0 is proportional to the brightness of the illumination and also to the total exposure of the photograph. Then $u(\mathbf{x}) = I_0 r(\mathbf{x})$.

The next step is to think of $r(\mathbf{x})$ as a member of a very large *ensemble* of scenes. We shall attempt a reconstruction of $r(\mathbf{x})$ given the data $\{\mathbf{X}_1 \ldots \mathbf{X}_N\}$. If r(x) were known, then the reconstruction problem would surely be trivial: there would be no need to look at the data. Since lack of knowledge is the essence of probability, we may think of $r(\mathbf{x})$ as a sample function of some stochastic process. Thus, our grainy photograph is generated by a pair of stochastic processes acting in series:

(i) Pick a scene $r(\mathbf{x})$ at random from the ensemble of scenes. Then choose I_0 (which we regard as deterministic) and set $u(\mathbf{x}) = I_0 r(\mathbf{x})$.

(ii) Generate $\{\mathbf{X}_k\}$, a Poisson process with density $u(\mathbf{x})$.

Consider the stochastic process that generates the scene $r(\mathbf{x})$. We assume that this process is *stationary*. This means that all statistical properties depend only on differences in the values of $\mathbf{x}$. It is important to remark that stationarity is a property of the ensemble and not of the individual scenes. For example, we can generate a stationary ensemble by starting with some complicated picture and then considering all possible translates of that picture. For our purposes, the only properties of the ensemble that matter are the first and second moments:

$$r_0 = E[r(\mathbf{x})], \tag{4}$$

$$\phi(\mathbf{x}' - \mathbf{x}) = E[(r(\mathbf{x}') - r_0)(r(\mathbf{x}) - r_0)]/r_0^2, \tag{5}$$

where E denotes the expected value $\equiv$ ensemble average. Thus r_0 is the mean reflectance and ϕ is the (normalized) correlation function. Note that r_0 and ϕ each have one fewer arguments than the expressions that define them. This is a consequence of stationarity.

It will be convenient in the following to work with $u(\mathbf{x})$ instead of $r(\mathbf{x})$. Therefore, we define

$$u_0 = I_0 r_0 = I_0 E[r(\mathbf{x})] = E[I_0 r(\mathbf{x})] = E[u(\mathbf{x})] \tag{6}$$

and note that ϕ can be written in the alternative form

$$\phi(\mathbf{x}' - \mathbf{x}) = E[(u(\mathbf{x}') - u_0)(u(\mathbf{x}) - u_0)]/u_0^2. \tag{7}$$

If the illumination (or the exposure) changes, u_0 changes proportionally, but ϕ remains invariant.

We are now ready to consider the reconstruction problem. Given $\{\mathbf{X}_k\}$, we seek a reconstruction of the image of the form

$$v(\mathbf{x}) = \sum_k h(\mathbf{x} - \mathbf{X}_k), \tag{8}$$

where $h(\mathbf{x})$ is to be determined. The interpretation of Equation 8 is that we take each photon and replace it by a function with shape h centered on the point where the photon was absorbed. The sum of all of these smeared-out photons is the reconstructed image. The reconstruction is *linear* because it satisfies the principle of superposition, it is *translation-invariant* or *stationary* because a translation of the photon image results in the corresponding translation of the reconstructed image. We call the operation described by Equation 8 a *filter* because $v(\mathbf{x})$ is the convolution of $h(\mathbf{x})$ with a sum of δ-functions centered at the points $\mathbf{X}_k$.

Our next task is to pick the function $h(\mathbf{x})$ in some optimal way. To do this, we need a definition of the error. First, define the contrast of the scene

$$C(\mathbf{x}) = \frac{r(\mathbf{x}) - r_0}{r_0} = \frac{u(\mathbf{x}) - u_0}{u_0}. \tag{9}$$

Note that $E[C(\mathbf{x})] = 0$ and also that $C(\mathbf{x})$ is independent of the brightness of the illumination of the scene. Let

$$v_0 = E[v(\mathbf{x})], \tag{10}$$

$$V(\mathbf{x}) = v(\mathbf{x}) - v_0. \tag{11}$$

The quantity $V(\mathbf{x})$ is essentially the contrast of the reconstructed scene. The reason that we do not divide by v_0 will become clear below.

Finally, we say that the *error* at the point $\mathbf{x}$ is

$$e(\mathbf{x}) = C(\mathbf{x}) - V(\mathbf{x}), \tag{12}$$

and we pick h to minimize $E[e^2(\mathbf{x})]$. (By stationarity, $E[e^2(\mathbf{x})]$ is, in fact, independent of $\mathbf{x}$.) If we divided by v_0 in the definition of V, then the error would be independent of the scale of v and hence independent of the scale of h. In this situation it would be impossible for there to be a unique h that minimizes $E[e^2(\mathbf{x})]$. Moreover, there is no need to divide by v_0 because the scale of v is fixed by the requirement that the error be small. Thus, v is always of order 1, independent of the scale of u. (A third reason not to divide by v_0 is that there are cases in which $v_0 = 0$!)

The next step is to evaluate $E[e^2(\mathbf{x})]$. The basic method for doing this is to use the following identity

$$E[\cdot] = E[E[E[\cdot \,|\, N, u] \,|\, u]]. \tag{13}$$

That is, to compute the expected value (ensemble average) of any quantity, we first take the expectation with a fixed scene u and a fixed number of photons N. Then we take the expectation over N (which is distributed according to the Poisson distribution), and finally we take the expectation over the ensemble of scenes u. Here is the calculation. First,

$$E[C^2(\mathbf{x})] = \phi(0) \tag{14}$$

by definition of ϕ. Next evaluate v_0:

$$E[v(\mathbf{x})\,|\,N, u] = \sum_{k=1}^{N} E[h(\mathbf{x} - \mathbf{X}_k)\,|\,N, u]$$
$$= N\mu^{-1} \int h(\mathbf{x} - \mathbf{X})u(\mathbf{X})d\mathbf{X}. \quad (15)$$

Then, since $E[N\,|\,u] = \mu$,

$$E[v(\mathbf{x})\,|\,u] = \int h(\mathbf{x} - \mathbf{X})u(\mathbf{X})d\mathbf{X}, \quad (16)$$

$$v_0 = E[v(\mathbf{x})] = u_0 \int h(\mathbf{X})\mathrm{d}\mathbf{X}. \quad (17)$$

Next, evaluate $E[C(\mathbf{x})V(\mathbf{x})]$:

$$E[C(\mathbf{x})V(\mathbf{x})\,|\,u] = C(\mathbf{x})E[V(\mathbf{x})\,|\,u]$$
$$= C(\mathbf{x}) \int h(\mathbf{x} - \mathbf{X})(u(\mathbf{X}) - u_0)d\mathbf{X}$$
$$= u_0 \int h(\mathbf{x} - \mathbf{X})C(\mathbf{x})C(\mathbf{X})d\mathbf{X}, \quad (18)$$
$$E[C(\mathbf{x})V(\mathbf{x})] = u_0 \int h(\mathbf{x} - \mathbf{X})\phi(\mathbf{x} - \mathbf{X})d\mathbf{X}$$
$$= u_0 \int h(\mathbf{X})\phi(\mathbf{X})d\mathbf{X} \quad (19)$$

since $\phi(\mathbf{X} - \mathbf{x}) = E[C(\mathbf{X})C(\mathbf{x})]$. Finally evaluate $E[V^2(\mathbf{x})]$:

$$E[v^2(\mathbf{x})\,|\,N, u] = \sum_{j,k=1}^{N} E[h(\mathbf{x} - \mathbf{X}_j)h(\mathbf{x} - \mathbf{X}_k)\,|\,N, u]$$
$$= N\mu^{-1} \int h^2(\mathbf{x} - \mathbf{X})u(\mathbf{X})d\mathbf{X}$$
$$+(N^2 - N)\mu^{-2} \int\int h(\mathbf{x} - \mathbf{X})h(\mathbf{x} - \mathbf{X}')u(\mathbf{X})u(\mathbf{X}')d\mathbf{X}\,d\mathbf{X}'. \quad (20)$$

In the last step, we made use of the fact that $\mathbf{X}_j$ and $\mathbf{X}_k$ are independent (for $j \neq k$) when N is given. Note that N is the number of diagonal terms and $N^2 - N$ is the number of off-diagonal terms in the double sum that appears in Equation 20. We now take the expectation over N.

Recall that the Poisson distribution has $E[N\,|\,u] = \mu$ and $E[N^2\,|\,u] = \mu^2 + \mu$. Therefore,

$$E[v^2(\mathbf{x})\,|\,u] = \int h^2(\mathbf{x} - \mathbf{X})u(\mathbf{X})d\mathbf{X}$$
$$+ \int\int h(\mathbf{x} - \mathbf{X})h(\mathbf{x} - \mathbf{X}')u(\mathbf{X})u(\mathbf{X}')d\mathbf{X}\mathrm{d}\mathbf{X}', \quad (21)$$

$$E[V^2(\mathbf{x})\,|\,u] = E[v^2(\mathbf{x})\,|\,u] - v_0^2$$
$$= \int h^2(\mathbf{x} - \mathbf{X})u(\mathbf{X})d\mathbf{X}$$
$$+ \int\int h(\mathbf{x} - \mathbf{X})h(\mathbf{x} - \mathbf{X}')(u(\mathbf{X})u(\mathbf{X}') - u_0^2)d\mathbf{X}d\mathbf{X}', \quad (22)$$

$$E[V^2(\mathbf{x})] = u_0 \int h^2(\mathbf{X})d\mathbf{X} + u_0^2 \int\int h(\mathbf{X})h(\mathbf{X}')\phi(\mathbf{X} - \mathbf{X}')d\mathbf{X}d\mathbf{X}'. \quad (23)$$

Combining the results of Equations 14, 19, and 23, we obtain

$$E[e^2(\mathbf{x})] = u_0 \int h^2(\mathbf{X})d\mathbf{X} + u_0^2 \int\int h(\mathbf{X})h(\mathbf{X}')\phi(\mathbf{x} - \mathbf{X}')d\mathbf{X}d\mathbf{X}' - 2u_0 \int h(\mathbf{X})\phi(\mathbf{X})d\mathbf{X} + \phi(0). \quad (24)$$

To calculate the optimal h is now an exercise in the calculus of variations. Let $h = h_0 + \epsilon h_1$, where $h_0(\mathbf{x})$ is the optimal filter and $h_1(\mathbf{x})$ is an arbitrary perturbation. Since h_0 minimizes $E[e^2(\mathbf{x})]$, we require that

$$0 = \frac{d}{d\epsilon} E[e^2(\mathbf{x})] = 2u_0 \int h_1(\mathbf{X}) \{h_0(\mathbf{X}) + u_0 \int h_0(\mathbf{X}')\phi(\mathbf{X} - \mathbf{X}')d\mathbf{X}' - \phi(\mathbf{X})\}d\mathbf{X}. \quad (25)$$

Since $h_1(\mathbf{X})$ is arbitrary, the quantity in { } must be zero. This gives the integral equation

$$h_0(\mathbf{X}) + u_0 \int h_0(\mathbf{X}')\phi(\mathbf{X} - \mathbf{X}')\, d\mathbf{X}' = \phi(\mathbf{X}), \quad (26)$$

which may be solved by taking Fourier transforms. Let

$$H_0(\boldsymbol{\xi}) = \int e^{-i\boldsymbol{\xi}\cdot\mathbf{x}}\, h_0(\mathbf{x})d\mathbf{x} \quad (27)$$

and similarly for $\Phi(\boldsymbol{\xi})$.

Then, since the Fourier transform converts convolutions into products,

$$H_0(\boldsymbol{\xi}) + u_0 H_0(\boldsymbol{\xi})\Phi(\boldsymbol{\xi}) = \Phi(\boldsymbol{\xi}), \quad (28)$$

$$H_0(\boldsymbol{\xi}) = \frac{\Phi(\boldsymbol{\xi})}{1 + u_0\Phi(\boldsymbol{\xi})}. \quad (29)$$

It should be mentioned that this result is the same as the optimal filter for extracting a signal with power spectrum Φ from an additive white noise. In our problem, however, the input to the filter is certainly not the *sum* of a signal and a noise. In fact, the signal is carried exclusively by the same (photon) events that generate the noise.

The low-light and high-light limits of Equation 29 are very instructive. As $u_0 \rightarrow 0$, $H_0 \rightarrow \Phi$. It follows, of course, that $h_0(\mathbf{x}) \rightarrow \phi(\mathbf{x})$. Thus, under low-light conditions, the best strategy is to spread each photon out according to the correlation function of the ensemble of scenes. Further spreading than this is never desirable, no matter how dim the light gets. As $u_0 \rightarrow \infty$, $H_0 \rightarrow 0$, but $u_0 H_0 \rightarrow 1$. This implies that $u_0 h(\mathbf{x}) \rightarrow \delta(\mathbf{x})$, the two-dimensional Dirac δ-function. Thus, under high-light conditions, the optimal filter reduces to a simple scaling of intensity with no spatial smoothing.

Strictly speaking, the high-light limit is slightly more complicated than the brief remarks in the foregoing paragraph implied. First, the statement that $u_0 H_0 \rightarrow 1$ is only correct at those frequencies $\boldsymbol{\xi}$ where $\Phi(\boldsymbol{\xi}) \neq 0$. Frequencies at which $\Phi = 0$ are always filtered out by the optimal filter, regardless of the light level. Second, even if $\Phi(\boldsymbol{\xi}) > 0$ for all $\boldsymbol{\xi}$, it is usually the case that $\Phi(\boldsymbol{\xi}) \rightarrow 0$ as $|\boldsymbol{\xi}| \rightarrow \infty$. In this case it is true that $u_0 H_0(\boldsymbol{\xi}) \rightarrow 1$ for each $\boldsymbol{\xi}$, but the limit is not uniform in $\boldsymbol{\xi}$. At any fixed u_0, therefore, we will still have $H_0(\boldsymbol{\xi}) \rightarrow 0$ as $|\boldsymbol{\xi}| \rightarrow \infty$, and the optimal filter has a "high frequency cutoff" that moves out to higher frequencies as u_0 increases.

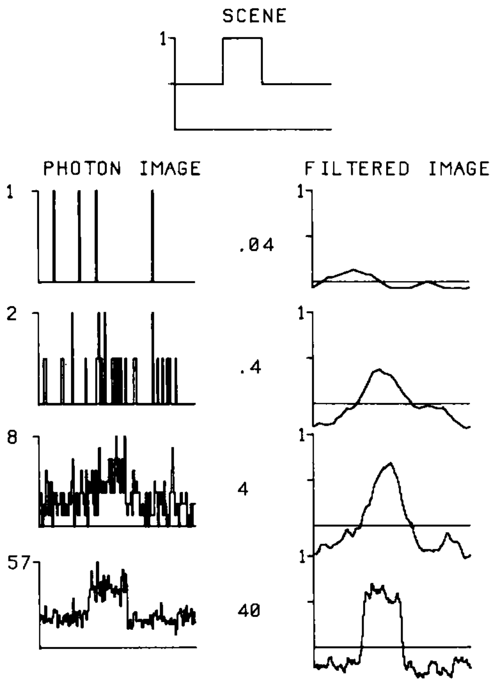


FIGURE 1. Simulation of photon noise and optimal filter: square pulse. Upper frame (SCENE) is the graph of the reflectance of a sample scene chosen at random from an ensemble of scenes. (See text for a precise description of the ensemble.) Left-hand column (PHOTON IMAGE) shows simulated results of photographing this scene at exposures that increase by factors of 10. (The numbers in the center column, .04, .4, 4, 40, give the exposure in arbitrary units.) The photon image is plotted as a histogram with 128 bins; note variable scale. Right-hand column (FILTERED IMAGE) shows the results of applying the optimal filter to the photon image. In the filtered image, the ordinate is the reconstructed *contrast* of the scene (see Equation 9). The optimal filter (not shown) is different for each exposure; it becomes sharper as the exposure is increased. The design of the optimal filter is based upon the statistical properties of the ensemble of scenes, but not upon the particular scene being photographed.

We conclude this section with some examples generated by computer simulation. The examples are one-dimensional and the space is periodic with period 2π. Therefore the relevant Fourier transform is the Fourier series. We construct a stationary ensemble of scenes in two different ways. In the first case, let

$$u(x, \theta) = u_1(1 + f(x - \theta)), \tag{30}$$

where u_1 is a constant, f is a given periodic function with period 2π, and where θ is a random variable uniformly distributed on the interval $(0, 2\pi)$. In this case the entire ensemble is characterized by the single random variable θ, and the symbol E takes on the concrete meaning

$$E[\cdot] = \frac{1}{2\pi}\int_0^{2\pi} (\cdot)d\theta. \tag{31}$$

Let $f(x)$ be represented by the Fourier series

$$f(x) = \sum_{k=-\infty}^{\infty} F(k)e^{ikx} \tag{32}$$

and similarly for $\phi(x)$ and $\Phi(k)$. We leave it as an exercise for the reader to show that

$$u_0 = u_1(1 + F(0)), \tag{33}$$

$$\Phi(0) = 0, \tag{34}$$

$$\Phi(k) = \frac{u_1^2}{u_0^2}|F(k)|^2, \qquad k \neq 0. \tag{35}$$

Thus the correlation function is easily expressed in terms of f, which defines the ensemble. Substituting (**34–35**) into (**29**) we obtain a formula for the optimal filter.

FIGURES 1, 2 and 3 show some results for this type of ensemble. In FIGURE 1, f is a square pulse; in FIGURE 2, f is a sinusoidal pulse; and in FIGURE 3, f is simply a sine wave. In all three figures the upper frame shows the function f with its random phase superimposed on the background (u_1 = .5). The left-hand column shows simulated "photographs" of this scene at different light levels increasing by factors of 10 (the

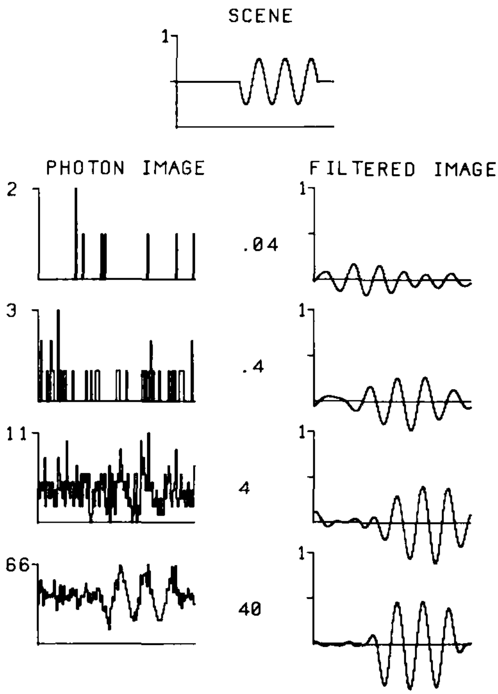


FIGURE 2. Simulation of photon noise and optimal filter: sinusoidal pulse. See legend of FIGURE 1.

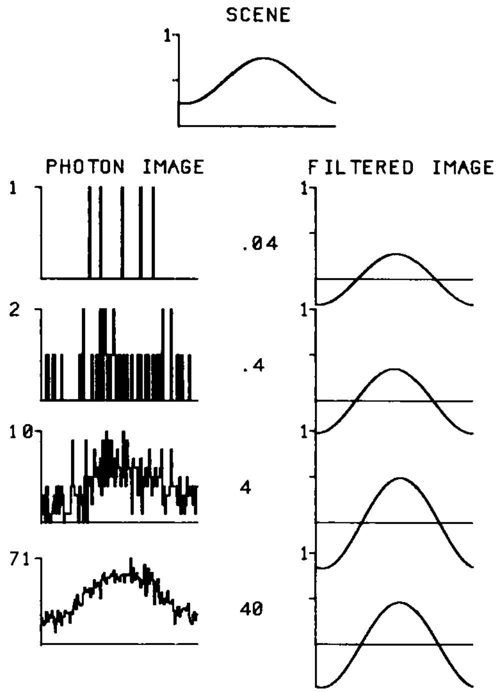


FIGURE 3. Simulation of photon noise and optimal filter: sine wave. See legend of FIGURE 1.

relative brightnesses are shown in the middle column). These photon images are presented as histograms with 128 bins; the plot shows the number of photons that happened to fall in each bin (note variable scale). For each light level an optimal filter is constructed on the basis of the correlation function of the ensemble of scenes. This optimal filter is applied to the photon image and the results are presented in the right-hand column. (Recall that the filtered image is an approximation to the *contrast* in the scene. As such, it has mean zero by definition, and its scale is also different from that of the scene, except in the special case when $u_0 = 1$.)

The filtered images in FIGURE 1 clearly show the gradual sharpening of the optimal filter as the light level is raised. Note that a clear-cut peak in the right place is already present at the light level .4; such a peak is not easy to discern in the photon image at this brightness. In FIGURE 2, a fair approximation to the sinusoidal pulse also emerges at the level .4, and, in FIGURE 3, the filtered image is spectacular at all light levels.

FIGURE 4 shows similar results for an ensemble constructed in a different way. The scene is constructed as the real part of a complex Fourier series in which the constant term (background) is given and all other coefficients are independent Gaussian random variables with mean zero and specified variance. The variance may be different for different coefficients. In the particular case shown in FIGURE 4 only a few coefficients are allowed to be different from zero. This is a more difficult case in the sense that the ensemble of scenes is characterized by more than just a single random variable.

VISION

In the case of vision, the situation differs from the foregoing in that time is an important variable. In this section, we consider the design of an optimal filter for smoothing the photon noise under conditions in which the scene varies in space and time. Thus, the scene is represented by a function $u(\mathbf{x}, t)$, which gives the expected number of photons per unit area per unit time, and the capture of a photon is an *event*: $(\mathbf{X}_k, T_k)$. The attempted reconstruction of the scene takes the form

$$v(\mathbf{x}, t) = \sum_k h(\mathbf{x} - \mathbf{X}_k, t - T_k), \tag{36}$$

where the function $h(\mathbf{x}, t)$ now satisfies the important constraint of *causality*

$$h(\mathbf{x}, t) = 0 \qquad \text{for } t < 0, \tag{37}$$

which means that the response to a photon cannot occur before the photon arrives. Equation 36 is a direct generalization of (8), but the new feature here is the causality constraint (37). As we shall see, this constraint makes it much more difficult to solve for the optimal filter.

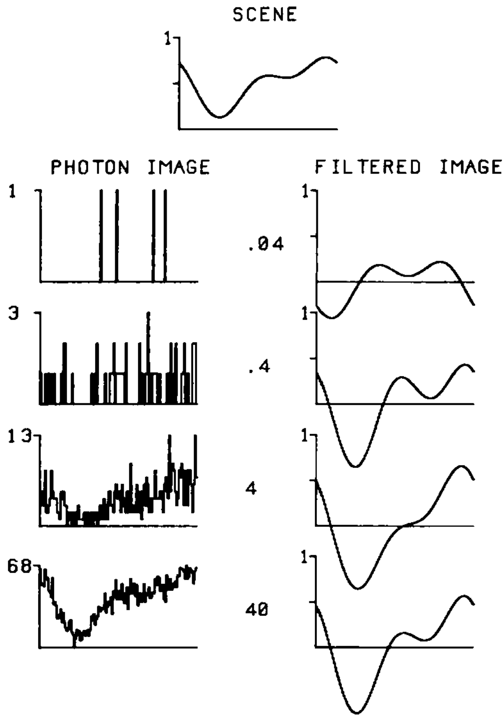


FIGURE 4. Simulation of photon noise and optimal filter: superposition of a few harmonics. See legend of FIGURE 1.

Up to and including (25), the theory in the space-time case is exactly the same as the theory for the purely spatial case that was developed in the foregoing section. Just replace x everywhere by (x, t) to obtain the corresponding result. Thus (25) becomes

$$0 = \int\int h_1(\mathbf{X}, T)\{h_0(\mathbf{X}, T) + u_0 \int\int h_0(\mathbf{X}', T')\phi(\mathbf{X} - \mathbf{X}', T - T')d\mathbf{X}'dT' - \phi(\mathbf{X}, T)\}d\mathbf{X}dT. \quad (38)$$

It is at this point that the constraint of causality plays a critical role. The functions h_0 and h_1 are both causal functions (zero for negative time). Therefore, h_1 is not arbitrary, and the most we can conclude is that the quantity in { } = 0 for $t > 0$. We express this by writing

$$h_0(\mathbf{X}, T) + u_0 \int\int h_0(\mathbf{X}', T')\phi(\mathbf{X} - \mathbf{X}', T - T')d\mathbf{X}'dT' - \phi(\mathbf{X}, T) = \psi(\mathbf{X}, T), \quad (39)$$

where $\psi(\mathbf{X}, T)$ is an *anti*causal function (for each **X**):

$$\psi(\mathbf{X}, T) = 0 \qquad \text{for } T > 0. \quad (40)$$

We now need Fourier transforms with respect to space and time. As before, let capital letters denote the spatial Fourier transform

$$H_0(\boldsymbol{\xi}, t) = \int e^{-i\boldsymbol{\xi}\cdot\mathbf{x}} h_0(\mathbf{x}, t)d\mathbf{x} \quad (41)$$

and use the notation ˆ for the temporal Fourier transform:

$$\hat{H}_0(\boldsymbol{\xi}, \omega) = \int e^{-i\omega t} H_0(\boldsymbol{\xi}, t)dt. \quad (42)$$

Applying both transformations to (39), we obtain

$$\hat{H}_0(\boldsymbol{\xi}, \omega) + u_0\hat{H}_0(\boldsymbol{\xi}, \omega)\hat{\Phi}(\boldsymbol{\xi}, \omega) - \hat{\Phi}(\boldsymbol{\xi}, \omega) = \hat{\Psi}(\boldsymbol{\xi}, \omega), \quad (43)$$

and applying the spatial transformation only to the causality conditions (37) and (40) we obtain

$$H_0(\boldsymbol{\xi}, t) = 0 \qquad \text{for } t < 0, \quad (44)$$

$$\Psi(\boldsymbol{\xi}, t) = 0 \qquad \text{for } t > 0. \quad (45)$$

As we shall see, the system (43–45) determines the optimal filter $\hat{H}_0$. Note that $\boldsymbol{\xi}$ is merely a parameter in this system. That is, we have a separate system for each $\boldsymbol{\xi}$. This is not true of ω because the causality conditions couple together the different temporal frequencies. In the following, we focus attention on ω and t, and all the statements that we make should be understood to hold for each $\boldsymbol{\xi}$.

The system (43–45) can be solved by the Wiener-Hopf technique of spectrum factorization,[3-5] which takes a particularly simple form in the present case. The key to this technique is the correspondence established by the Fourier transform between functions of time that are causal and functions of complex ω that are bounded and analytic in the lower-half plane. The first step in the Wiener-Hopf technique is the factorization of the spectrum $1 + u_0\hat{\Phi}$ into the product of two functions, $\hat{A}^{(+)}$ and $\hat{A}^{(-)}$, such that $\hat{A}^{(+)}$ and $1/\hat{A}^{(+)}$ are both the transforms of causal functions, while

$\hat{A}^{(-)}$ and $1/\hat{A}^{(-)}$ are both the transforms of anticausal functions. The factorization can sometimes be achieved by inspection (when Φ is a rational function), but the following algorithm works nicely even when Φ is only given in the form of numerical data for real ω.

Let

$$\hat{B} = \log(1 + u_0\hat{\Phi}). \tag{46}$$

For real ω, $\hat{\Phi}$ is a real, positive, even function of ω, so there is no difficulty with this definition of $\hat{B}$, and $\hat{B}$ is also an even function. Moreover, in typical cases, $\hat{\Phi} \to 0$ as $\omega \to \infty$, so $\hat{B}$ inherits this property. The (temporal) inverse transform of $\hat{B}$ is a time function $B(\xi, t)$, and we partition this function into its causal and anticausal parts:

$$B(\xi, t) = B_+(\xi, t) + B_-(\xi, t) \tag{47}$$

where B_+ is a causal function and B_- an anticausal function of t. Taking Fourier transforms of B_+ and B_- and exponentiating, we get the required factorization:

$$1 + u_0\Phi = \exp(\hat{B}) = \exp(\hat{B}_+)\exp(\hat{B}_-) = \hat{A}^{(+)}\hat{A}^{(-)}, \tag{48}$$

where

$$\hat{A}^{(+)} = \exp(\hat{B}_+), \tag{49}$$

$$\hat{A}^{(-)} = \exp(\hat{B}_-). \tag{50}$$

As functions of complex ω, $\hat{A}^{(+)}$ and its reciprocal $1/\hat{A}^{(+)}$ are bounded and analytic in the lower-half plane. This follows from the foregoing construction, since B_+ is a causal function, since exponentiation is analytic everywhere, and since $|\exp(z)| \le \exp|z|$. The corresponding argument shows that $\hat{A}^{(-)}$ and $1/\hat{A}^{(-)}$ are bounded and analytic in the upper-half plane.

Having obtained the factorization, we are now ready to consider Equation 43, which we rewrite as follows:

$$(1 + u_0\hat{\Phi})(\hat{H}_0 - u_0^{-1}) = \hat{\Psi} - u_0^{-1}. \tag{51}$$

Dividing by $\hat{A}^{(-)}$, we obtain

$$\hat{A}^{(+)}(\hat{H}_0 - u_0^{-1}) = (\hat{\Psi} - u_0^{-1})/\hat{A}^{(-)}. \tag{52}$$

This equation holds for real ω. But the function on the left (right) can be extended as a bounded, analytic function in the lower- (upper-) half plane. Let $F(\omega)$ be the function on the left for ω in the lower-half plane and let $F(\omega)$ be the function on the right for ω in the upper-half plane. Then F is analytic and bounded in each half plane and it is continuous on the real axis. The only function with these properties is a constant. Therefore

$$\hat{A}^{(+)}(\hat{H}_0 - u_0^{-1}) = C. \tag{53}$$

We can evaluate C by letting $\omega \to \pm\infty$. From the construction of $\hat{A}^{(+)}$ it is clear that $\hat{A}^{(+)} \to 1$ as $\omega \to \pm\infty$. Moreover, it seems reasonable to impose the regularity condition $\hat{H}_0(\xi, \omega) \to 0$ as $\omega \to \pm\infty$. In that case, we have $C = -u_0^{-1}$ and

$$\hat{H}_0 = \frac{1}{u_0}\left(1 - \frac{1}{\hat{A}^{(+)}}\right). \tag{54}$$

This is better expressed if we let

$$\hat{A}^{(+)} = 1 + \hat{K}^{(+)} \tag{55}$$

so that $\hat{K}^{(+)} \to 0$ as $\omega \to \pm\infty$. Then

$$\hat{H}_0 = \frac{1}{u_0}\frac{\hat{K}^{(+)}}{1 + \hat{K}^{(+)}}. \tag{56}$$

This is the formula for the optimal filter. An important remark is that $\hat{K}^{(+)}$ depends on u_0, so the dependence on light level is not as simple as it appears to be in Equation 56.

As in the foregoing section we now consider the low-light and high-light limits. These are most easily read off from the equation

$$(1 + u_0\hat{\Phi})\hat{H}_0 - \hat{\Phi} = \hat{\Psi}. \tag{57}$$

When $u_0 = 0$, we have

$$\hat{H}_0 - \hat{\Phi} = \hat{\Psi} \tag{58}$$

or

$$\hat{H}_0 - \hat{\Phi}_+ = \hat{\Psi} + \hat{\Phi}_-, \tag{59}$$

where $\hat{\Phi}_+$ and $\hat{\Phi}_-$ are the Fourier transforms of the causal and anticausal parts of $\Phi(\xi, t)$. It follows that

$$\hat{H}_0 = \hat{\Phi}_+. \tag{60}$$

This should be compared with the low-light limit of the unconstrained filter, which was $H_0 = \Phi$.

When u_0 is large, on the other hand, we may neglect the 1 in (57) in comparison with $u_0\hat{\Phi}$. In that case, we have

$$(u_0\hat{H}_0 - 1)\Phi = \hat{\Psi}, \tag{61}$$

which is clearly solved by setting $u_0\hat{H}_0 = 1$ and $\hat{\Psi} = 0$. We conclude that

$$\lim_{u_0 \to \infty} u_0\hat{H}_0 = 1. \tag{62}$$

This is the same high-light limit as in the unconstrained case. The reader may want to verify that the low-light and high-light limits can also be obtained from the formula (56). This can be done by studying the behavior of the factorization for small and large values of u_0.

To illustrate the behavior of the optimal filter in more detail, we consider the design of a "retina" for looking at an ensemble of spots of light (or shade) that execute Brownian motion in the plane against a fixed background. We assume that the spot k makes its first appearance at $(\mathbf{X}_k^0, T_k^0)$; these events (not to be confused with photon events) are distributed as a Poisson process in space-time with uniform density ρ_0 (per unit area per unit time). Following its appearance, the kth spot of light executes a Brownian motion (continuous random walk) in the plane so that its trajectory is given by

$$\mathbf{X}_k(t) = \mathbf{X}_k^0 + \mathbf{Y}_k(t - T_k^0) \tag{63}$$

where $\mathbf{Y}_k(t)$ is an independently chosen Brownian motion path with $\mathbf{Y}_k(0) = 0$ and diffusion coefficient D. Spot k remains turned on for a random time T_k, which is independently chosen from the exponential density $\nu_0 \exp(-\nu_0 t)$. The ensemble of scenes may therefore be written

$$u(\mathbf{x}, t) = u_1\{1 + \sum_k f(\mathbf{x} - \mathbf{X}_k^0 - \mathbf{Y}_k(t - T_k^0))S(t - T_k^0)\, S(T_k^0 + T_k - t)\}, \quad (64)$$

where f is a fixed function describing the shape of the individual spots of light and where $S(t)$ is the unit step function ($S(t) = 1$ for $t > 0$, $S(t) = 0$ for $t < 0$). We leave it as a (fairly difficult) exercise for the reader to show that

$$u_0 = u_1\left(1 + \frac{\rho_0}{\nu_0} F(0)\right), \quad (65)$$

$$\hat{\Phi}(\xi, \omega) = \frac{2\alpha(\xi)\omega_0(\xi)}{\omega^2 + \omega_0^2(\xi)}, \quad (66)$$

where

$$\alpha(\xi) = \frac{\rho_0\nu_0^{-1}|F(\xi)|^2}{(1 + \rho_0\nu_0^{-1} F(0))^2}, \quad (67)$$

$$\omega_0(\xi) = \nu_0 + D|\xi|^2. \quad (68)$$

A crucial role in the derivation of (66) is played by a density function $\rho(\mathbf{x}, t)$, which satisfies the heat equation with leakage

$$\frac{\partial \rho}{\partial t} = D\Delta\rho - \nu_0\rho, \quad (69)$$

where Δ is the Laplace operator. The Fourier transform of this equation (in space and time) involves the operator

$$i\omega + D|\xi|^2 + \nu_0 = i\omega + \omega_0(\xi). \quad (70)$$

This explains the appearance of $\omega_0(\xi)$ in the formula for $\hat{\Phi}$.

The power spectrum $\hat{\Phi}$ derived above is a very simple rational function of ω, and the factorization of $1 + u_0\hat{\Phi}$ can be carried out by inspection:

$$\begin{aligned} 1 + u_0\hat{\Phi} &= \frac{\omega^2 + \omega_0^2 + 2u_0\,\alpha\,\omega_0}{\omega^2 + \omega_0^2} \\ &= \left(\frac{i\omega + (\omega_0^2 + 2u_0\alpha\omega_0)^{1/2}}{i\omega + \omega_0}\right)\left(\frac{-i\omega + (\omega_0^2 + 2u_0\alpha\omega_0)^{1/2}}{-i\omega + \omega_0}\right) \\ &= \hat{A}^{(+)}\hat{A}^{(-)}. \end{aligned} \quad (71)$$

Therefore

$$\hat{H}_0(\xi, \omega) = \frac{1}{u_0}\,\frac{\left(1 + \dfrac{2u_0\alpha(\xi)}{\omega_0(\xi)}\right)^{1/2} - 1}{i\dfrac{\omega}{\omega_0(\xi)} + \left(1 + \dfrac{2u_0\alpha(\xi)}{\omega_0(\xi)}\right)^{1/2}}. \quad (72)$$

We shall approximate this formula for the optimal filter in the range of light levels and spatial frequencies for which

$$\frac{2u_0\alpha(\boldsymbol{\xi})}{\omega_0(\boldsymbol{\xi})} \ll 1. \tag{73}$$

That is, we consider the low-light or high- (spatial) frequency behavior of the optimal filter. Making this approximation, we find

$$\begin{aligned}\hat{H}_0(\boldsymbol{\xi}, \omega) &= \frac{\alpha(\boldsymbol{\xi})}{(i\omega + |\boldsymbol{\xi}|^2 D + \nu_0) + u_0\alpha(\boldsymbol{\xi})}\\ &= \frac{\alpha(\boldsymbol{\xi})(i\omega + |\boldsymbol{\xi}|^2 D + \nu_0)^{-1}}{1 + u_0\alpha(\boldsymbol{\xi})(i\omega + |\boldsymbol{\xi}|^2 D + \nu_0)^{-1}}.\end{aligned} \tag{74}$$

Incidentally, although this expression was derived for low-light levels, it has the correct high-light limit: $u_0 H_0 \to 1$ as $u_0 \to \infty$.

Note that the approximate optimal filter given by (74) can be realized by the feedback network shown in FIGURE 5. In this figure the input is $\hat{I}$, the output is $\hat{O}$, and the difference signal is $\hat{E}$. The mean light level u_0 appears as a simple gain (independent of frequency) in the feedback path. All other elements are independent of the mean light level. The box $\alpha(\boldsymbol{\xi})$ represents a static network with no dynamics. In fact, in the special case where the spots of light in the scene are concentrated at individual points, $f(\mathbf{x}) = \delta(\mathbf{x})$, and $F(\boldsymbol{\xi}) = F(0) = 1$, so α is independent of $\boldsymbol{\xi}$. The box $(i\omega + |\boldsymbol{\xi}|^2 D + \nu_0)^{-1}$ computes the solution of the diffusion equation with leakage. In neural terms it can be realized by a dendrite-like network (such as the horizontal cell network) with current input and voltage output. The operation of the feedback loop can be understood as follows. Under low-light conditions the feedback is negligible, and the receptive field is determined by the solution of the diffusion equation. As the light level is raised, the feedback becomes more and more important. This has the effect of driving $\hat{E}$ towards zero, which means that the approximation $\hat{I} \approx u_0\hat{O}$ becomes better at higher levels of light. This has the effect of sharpening and speeding up the response to a photon. *Note that these changes in the overall spatial and temporal dynamics of the receptive field are achieved merely by altering a single feedback gain; there is no selective adjustment of nearby versus faraway connections nor is there any resetting of time constants of the individual components of the system.*

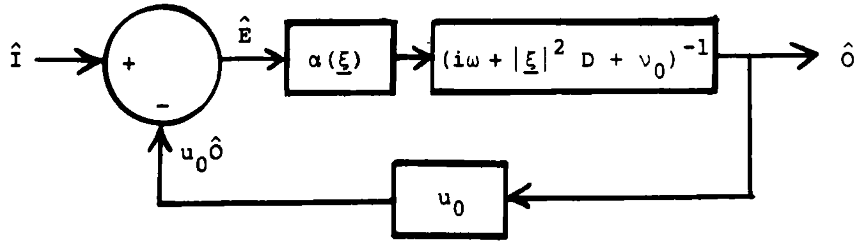
Does the retina in fact function as an optimal filter? This is an active area of current research for us. Tranchina *et al.*[6] have measured temporal frequency transfer functions of horizontal cells in the turtle retina. The stimulus was spatially uniform light with time dependence $u_0 + a_0 \cos(2\pi f t)$, where u_0 is the mean light level, a_0 is the amplitude of the sinusoidal component, and f is the frequency. The response was of the form $u_1 + a_1 \cos(2\pi f t + \theta)$. The temporal transfer function $H(f)$ is defined as $H(f) = (a_1/a_0)\exp(i\theta)$. It was found that a family of temporal transfer functions measured over a large range of mean light levels could be fit by a feedback model in which the gain of feedback is proportional to the mean light level.[6] The transfer function of such a feedback model is of the form

$$H(f) = \frac{M(f)}{1 + u_0 L(f) M(f)}, \tag{75}$$

where $M(f)$ is the low-light limit of H and where $L^{-1}(f)$ is the high-light limit of $u_0 H$. Note that Equation 75 can be realized by the feedback network shown in FIGURE

6. The resemblance to the feedback network of FIGURE 5 is striking. The most important difference is the presence of the nontrivial network $L(f)$ in the feedback pathway. This is a consequence of the fact that the observed high-light limit is not flat, but tends to emphasize the higher frequencies.

To incorporate such behavior into the optimal filtering theory requires a generalization of the simple form of the theory presented in the foregoing. That is, we have to assume that the retina is designed not to reconstruct the contrast of the original scene



$$\hat{E} = \hat{I} - u_0\hat{O}$$

$$\hat{O} = \frac{\alpha(\underline{\xi})}{i\omega + |\underline{\xi}|^2 D + \nu_0}$$

Therefore:

$$\hat{O} = \frac{\alpha(\underline{\xi})(i\omega + |\underline{\xi}|^2 D + \nu_0)^{-1}}{1 + u_0\,\alpha(\underline{\xi})(i\omega + |\underline{\xi}|^2 D + \nu_0)^{-1}}\,\hat{I}$$

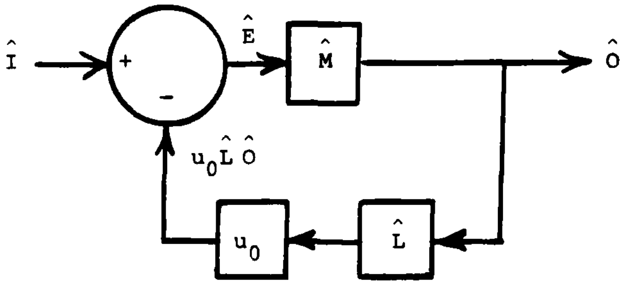
FIGURE 5. Feedback realization of the (approximate) optimal spatiotemporal filter for looking at a collection of spots of light that: (i) turn on at random points in space-time, (ii) move according to Brownian motion paths with diffusion coefficient D, and (iii) turn off at random times with ν_0 being the probability of extinction per unit time. $\hat{I}$ = input, $\hat{E}$ = difference signal, $\hat{O}$ = output. The box labeled $\alpha(\xi)$ is a static (that is, instantaneous) network with no dynamics. The box labeled $(i\omega + |\xi|^2 D + \nu_0)^{-1}$ is a dendritic network such as the horizontal cell layer of the vertebrate retina. The feedback path contains a simple gain proportional to the mean light level u_0. In this model retina, light adaptation is achieved entirely by changing this one parameter. Note that the effects of u_0 on the overall spatiotemporal transfer function are complicated. In particular, the response of the network to a photon becomes faster and sharper as u_0 is increased.

per se, but rather to construct a filtered version of the contrast of the original scene. We are now working on a quantitative test of the generalized theory.

NUCLEAR MEDICINE

We now consider the problem of designing an optimal linear filter for removing the photon (or other particle) noise from a nuclear medicine scan. There are several differences between the nuclear medicine case and the problems that we have considered up to now.

First, the ensemble of "scenes" or photon densities now becomes a concrete reality: it is a record of the scans of all (or a selected group of) patients who have received a similar nuclear medicine scan in the past. The definition of "similar" raises interesting problems that are outside the scope of this paper. Another difficulty with the design of a filter based on an ensemble of patient scans is that the scans are themselves noisy; this does not turn out to be a major obstacle, however. A third objection may be the difficulty of storing and retrieving the large amounts of data involved, but these difficulties are being rapidly overcome by progress in the development of computers. Perhaps the most serious question is whether it makes sense to use information from



$$\hat{E} = \hat{I} - u_0\hat{L}\hat{O}$$

$$\hat{O} = \hat{M}\hat{E}$$

Therefore:

$$\hat{O} = \frac{\hat{M}}{1 + u_0\hat{L}\hat{M}}\hat{I}$$

FIGURE 6. Feedback model of light adaptation in the turtle retina.[6] $\hat{I}$ = input, $\hat{E}$ = difference signal, $\hat{O}$ = output. $\hat{M}(f)$ = forward gain at frequency f, $u_0\hat{L}(f)$ = gain of the feedback path, where u_0 = mean light level. When u_0 is small, there is essentially no feedback and the transfer function is approximately $\hat{M}$. When u_0 is large, the difference signal is small, so that we have approximately $\hat{I} = u_0\hat{L}\hat{O}$. Compare FIGURE 5.

previous patients to process the picture of a given patient. This idea may seem less radical when we recall that this is exactly what a doctor does when he makes a diagnosis. The symptoms are not considered in isolation; they are evaluated in the context of the *a priori* probabilities (sometimes called the "index of suspicion") that the patient may have a particular disease. These *a priori* probabilities are learned from "experience." This means that the physician has in the back (maybe in the front) of his mind a record of the ensemble of patients that he has either seen himself or heard of through the medical literature. Every physician makes extensive use of such information in arriving at a diagnosis. Here we apply the same idea in a quantitative way.

The second new feature in the nuclear medicine case is that the patient ensemble is certainly not stationary. This is because the nuclear medicine camera is always being pointed *at* something. The organ in question is centered in the image, which therefore has a natural origin. This means that the Fourier transform, our main tool up to now, will no longer be useful. It also implies that the ensemble has to be characterized in a more complicated way. The expectation over the ensemble is now a function of position and the correlation function is now genuinely a function of two variables, not just of their difference. For the same reason, the idea of "contrast" is no longer useful and the mean values cannot be thrown away. Our goal will be to reconstruct the true photon densities. The lack of stationarity is the major mathematical difficulty in the nuclear medicine case.

This difficulty is compensated by an advantage, however. Since we are free to process the nuclear medicine scan "off-line," the constraint of causality does not apply even when time is an important variable, as in nuclear medicine scans of the heart. That is, once the study is complete and the data have been stored, there is no reason why the filter cannot look both forward and backward in time. Thus, time and space play identical roles here and we shall not even bother to distinguish between them.

Finally, there is the following minor difference: a nuclear medicine scan is usually recorded in discrete cells called pixels. Thus the data take the form of the number of counts in each pixel and not the form of a list of the coordinates of the individual photon events. We shall develop the theory in this section accordingly, but we emphasize that this is not a fundamental issue, since the discrete and continuous versions of the theory (while they look somewhat different) are essentially the same.

Let the pixels in a nuclear medicine scan be designated by an index $p = 1, 2, \ldots, P$. The order is of no significance and we use this notation even when the study involves time. For example, pixels $1 \ldots p_1$ might be the first frame in time, $p_1 + 1 \ldots 2p_1$ the second frame, and so on. Let N_p be the number of counts recorded in pixel p. For a given patient this is a random variable distributed according to the Poisson distribution

$$Pr\{N_p = n\} = \frac{U_p^n}{n!} e^{-U_p}, \tag{76}$$

where U_p is the expected number of counts in pixel p for the patient in question. Note that $(U_1 \ldots U_P)$ are random variables only because we regard the selection of a patient from the ensemble as a random process; for any particular patient $(U_1 \ldots U_P)$ have definite numerical values. These values are, however, unknown, and the point of performing a nuclear medicine scan is to determine them (approximately). For a given patient, the random variables N_p and N_q are independent, for $p \neq q$. Note the important point that N_p and N_q are only *conditionally* independent, the condition being that the expected values U_p and U_q are known. Once these expectations are known, we cannot obtain any further information about N_p by measuring N_q. It could happen, though, that when we consider the ensemble as a whole, U_p and U_q are strongly correlated. In that case, if we do not know U_p and U_q, then a measurement of N_p tells us a lot about N_q.

Since the individual patient is characterized by the list of numbers $U_1 \ldots U_P$, a complete description of the ensemble would be a statement of the joint probability distribution function

$$F(u_1 \cdots u_P) = Pr\{U_1 \leq u_1 \text{ and } \cdots \text{ and } U_P \leq u_P\}. \tag{77}$$

Fortunately, we have no need for the function F per se, but we only need the ensemble average and the correlation function:

$$\bar{U}_p = E[U_p], \tag{78}$$

$$\phi_{pq} = E[(U_p - \bar{U}_p)(U_q - \bar{U}_q)]. \tag{79}$$

Note that the correlation function here is not normalized. Also, $\bar{U}_p$ depends on p and ϕ_{pq} is not merely a function of $p - q$. The problem of how to *measure* $\bar{U}_p$ and ϕ_{pq} will be considered below. For now we regard them as given functions.

In summary, the data N_p are generated in a two-step stochastic process. The first step is the selection of a patient $U_1 \ldots U_P$ from the ensemble. Then, given $U_1 \ldots U_P$, the counts $N_1 \ldots N_P$ are generated independently in each pixel using the Poisson distribution with mean values $U_1 \ldots U_P$, respectively.

Our goal is an optimal linear reconstruction of $U_1 \ldots U_P$ from the data $N_1 \ldots N_P$. That is, we seek a reconstruction of the form

$$V_p = \bar{U}_p + \sum_{q=1}^{P} H_{pq}(N_q - \bar{U}_q). \tag{80}$$

The first term in (80) is just the ensemble average. It would be our best guess in the absence of data. The second term is a correction to the ensemble average in the form of an arbitrary linear combination of the *deviations* from the ensemble average in the different pixels. The function H_{pq} is the (nonstationary) filter; it remains to be determined. The error in pixel p is simply $V_p - U_p$ (*not* $V_p - \bar{U}_p$); therefore we seek H_{pq} to minimize $E[(V_p - U_p)^2]$. Note that this is a separate problem for each p; we minimize the error in each pixel separately with no trade-off between different pixels. It will turn out, though, that all of the different problems are linear systems of equations with the same matrix (but different right-hand sides).

We now proceed (as in the foregoing sections) to evaluate the mean square error and then choose H to minimize the resulting expression. It is convenient to start in the following way:

$$\begin{aligned} E[(V_p - U_p)^2] &= E[\{(V_p - \bar{U}_p) - (U_p - \bar{U}_p)\}^2] \\ &= E[(V_p - \bar{U}_p)^2] - 2E[(V_p - \bar{U}_p)(U_p - \bar{U}_p)] \\ &\quad + E[(U_p - \bar{U}_p)^2]. \end{aligned} \tag{81}$$

The last term is just ϕ_{pp}, by definition. To evaluate the other two terms we first take the expectation with $U = (U_1 \ldots U_P)$ given; then we take the expectation over the patient ensemble:

$$E[(V_p - \bar{U}_p) \mid U] = \sum_q H_{pq}(U_q - \bar{U}_q), \tag{82}$$

$$E[(V_p - \bar{U}_p)(U_p - \bar{U}_p)] = \sum_q H_{pq}\phi_{pq}, \tag{83}$$

$$E[(V_p - \bar{U}_p)^2 \mid U] = \sum_{q,q'} H_{pq}H_{pq'}\, E[(N_q - \bar{U}_q)\,(N_{q'} - \bar{U}_{q'}) \mid U]. \tag{84}$$

This last expression must be evaluated with care. For $q \neq q'$, N_q and $N_{q'}$ are independent (given U). Therefore

$$E[(N_q - \bar{U}_q)(N_{q'} - \bar{U}_{q'}) \mid U] = (U_q - \bar{U}_q)(U_{q'} - \bar{U}_{q'}), \qquad q \neq q'. \tag{85}$$

On the other hand, for $q = q'$ we get

$$E[(N_q - \bar{U}_q)^2 | U] = E[\{(N_q - U_q) + (U_q - \bar{U}_q)\}^2 | U]$$
$$= E[(N_q - U_q)^2 | U] + E[(N_q - U_q) | U](U_q - \bar{U}_q) + (U_q - \bar{U}_q)^2$$
$$= U_q + (U_q - \bar{U}_q)^2, \tag{86}$$

where we have used the important property of the Poisson distribution that the variance is equal to the mean. Thus, when $q = q'$ the extra term U_q appears. It follows that

$$E[(V_p - \bar{U}_p)^2 | U] = \sum_q H_{pq}^2 U_q + \sum_{q,q'} H_{pq} H_{pq'} (U_q - \bar{U}_q)(U_{q'} - \bar{U}_{q'}), \tag{87}$$

$$E[(V_p - \bar{U}_p)] = \sum_q H_{pq}^2 \bar{U}_q + \sum_{qq'} H_{pq} H_{pq'} \phi_{qq'}. \tag{88}$$

Substituting Equations 83 and 88 into Equation 81, we get the required formula:

$$E[(V_p - U_p)^2] = \sum_q H_{pq}^2 \bar{U}_q + \sum_{qq'} H_{pq} H_{pq'} \phi_{qq'} - 2 \sum_q H_{pq} \phi_{pq} + \phi_{pp}. \tag{89}$$

To find the optimal H_{pq} set $H_{pq} = H_{pq}^0 + \epsilon H_{pq}^1$, differentiate with respect to ϵ, and set the result equal to zero for $\epsilon = 0$ and arbitrary H_{pq}^1. Alternatively (since the problem is discrete), just differentiate with respect to $H_{p\ell}$ and set the result equal to zero. In either case, we get the following linear system for H_{pq}:

$$H_{pq} \bar{U}_q + \sum_{q'} H_{pq'} \phi_{qq'} = \phi_{pq}. \tag{90}$$

For each p, this is a linear system of equations for the unknowns $(H_{p1} \ldots H_{pP})$. The matrix of the system is independent of p; its qq' element is $\bar{U}_q \delta_{qq'} + \phi_{qq'}$. It is easy to see that this is a symmetric, positive definite matrix. Symmetry is a direct consequence of the definition of ϕ, and positivity is proved as follows. Let X_q be an arbitrary vector (not identically 0). Then

$$\sum_{qq'} X_q X_{q'} (\bar{U}_q \delta_{qq'} + \phi_{qq'}) = \sum_q \bar{U}_q X_q^2 + \sum_{qq'} X_q X_{q'} E[(U_q - \bar{U}_q)(U_{q'} - \bar{U}_{q'})]$$
$$= \sum_q \bar{U}_q X_q^2 + E\left[\left\{\sum_q X_q (U_q - \bar{U}_q)\right\}^2\right] > 0. \tag{91}$$

In practice it may not be convenient to store the matrix H_{pq}, since the number of elements in this matrix is the *square* of the number of pixels in the study. In practical cases the correlation function ϕ_{pq} may be sparse (contain mostly zero elements). This would happen if each pixel were correlated only with a limited number of other pixels. We could, for example, set $\phi_{pq} = 0$ when $|\phi_{pq}|$ is measured to be less than some predetermined tolerance. Nevertheless, the filter H_{pq} would usually be dense. We therefore consider alternative strategies to the direct computation and storage of H_{pq}.

First, we summarize the foregoing results in matrix-vector notation. Let $V = (V_1 \ldots V_P)$, $\bar{U} = (\bar{U}_1 \ldots \bar{U}_P)$, etc. Similarly, let H, ϕ be the matrices with elements H_{pq}

and ϕ_{pq} respectively. Also let $D(\overline{U})$ be the diagonal matrix with $\overline{U}_1 \ldots \overline{U}_P$ on the diagonal. Our reconstruction takes the form

$$V = \overline{U} + H(N - \overline{U}), \tag{92}$$

where H is the solution of

$$H(D(\overline{U}) + \phi) = \phi. \tag{93}$$

The idea now is to eliminate H from these equations. We do this by introducing the vector W, which is defined as the solution of

$$(D(\overline{U}) + \phi)W = N - \overline{U}. \tag{94}$$

Substituting this expression for $N - \overline{U}$ into (92) we obtain

$$\begin{aligned} V &= \overline{U} + H(D(\overline{U}) + \phi)W \\ &= \overline{U} + \phi W. \end{aligned} \tag{95}$$

Therefore, we can avoid any explicit reference to H by using the following algorithm to obtain the reconstruction V. Given data N on a particular patient, we solve the linear system (94) for W. Then we use the result in (95) to obtain V. With this approach we have to solve a linear system for each patient. If ϕ is sufficiently sparse and H is dense, however, the cost of *solving* for W may be less than the cost of merely multiplying H by $(N - \overline{U})$. Also, we avoid the difficulty of storing H.

We now turn to the important problem of determining $\overline{U}$ and ϕ from data in the form of a concrete ensemble of patient scans. Let the scan of the jth patient in the ensemble be denoted $(N_1^j \ldots N_P^j)$. Let n_e be the number of patients in the ensemble, so $j = 1 \ldots n_e$. We want to estimate $\overline{U}$ and ϕ, and the difficulty is that these quantities are defined in terms of $(U_1^j \ldots U_P^j)$, which are unknown!

Fortunately, this difficulty can be overcome. First we note that

$$\begin{aligned} E[N_p] &= E[E[N_p | U]] \\ &= E[U_p] = \overline{U}_p. \end{aligned} \tag{96}$$

This suggests the estimate

$$\overline{U}_p^{(\text{est})} = \frac{1}{n_e} \sum_{j=1}^{n_e} N_p^j, \tag{97}$$

which is the obvious way to estimate the expected number of counts in pixel p.

Next, we combine (85) and (86) to obtain

$$E[(N_p - \overline{U}_p)(N_q - \overline{U}_q) | U] = U_p \delta_{pq} + (U_p - \overline{U}_p)(U_q - \overline{U}_q), \tag{98}$$

from which it follows that

$$\begin{aligned} E[(N_p - \overline{U}_p)(N_q - \overline{U}_q)] &= \overline{U}_p \delta_{pq} + \phi_{pq} \\ &= (D(\overline{U}) + \phi)_{pq}. \end{aligned} \tag{99}$$

This suggests the estimate:

$$(D(\overline{U}) + \phi)_{pq}^{(\text{est})} = \frac{1}{n_e} \sum_{j=1}^{n_e} (N_p^j - \overline{U}_p^{(\text{est})})(N_q^j - \overline{U}_q^{(\text{est})}). \tag{100}$$

This result is more surprising, since the right-hand side looks like an estimate of ϕ alone. In fact, however, the additional diagonal terms emerge out of the fluctuations in the data! Another nice feature of this method for estimating $D(\overline{U}) + \phi$ is that the estimate is automatically symmetric and positive definite.

It is easy to see that $\overline{U}_p^{(\text{est})}$ is an *unbiased* estimate. That is

$$E[\overline{U}_p^{(\text{est})}] = \overline{U}_p. \quad (101)$$

If we apply the same test to the matrix estimate we find (after a calculation that is omitted here)

$$E[(D(\overline{U}) + \phi)_{pq}^{(\text{est})}] = (D(\overline{U}) + \phi)_{pq} - \frac{1}{n_e}\overline{U}_p\,\delta_{pq}. \quad (102)$$

Thus the diagonal terms are underestimated (on the average) by an amount which is $O(1/n_e)$. This can be easily fixed up (if desired) by adding $\overline{U}_p^{(\text{est})}/n_e$ to the diagonal terms of the matrix.

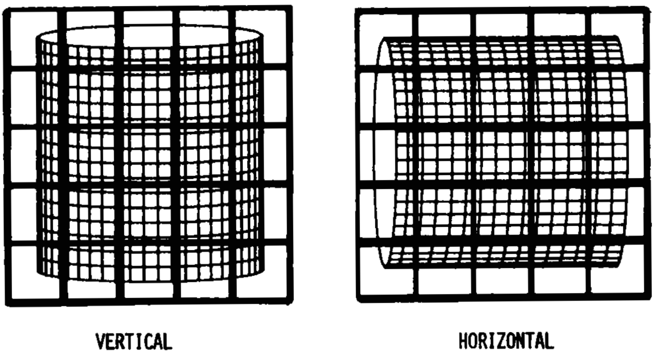


FIGURE 7. Vertical and horizontal cylinders for computer simulation of nuclear medicine scans. The cylinders are filled with a medium containing radioactive isotope. The 5 × 5 grid in the foreground shows the simulated camera pixels. The two cylinders have the same (square) projections onto this grid, but their scans will differ because the expected number of counts in a pixel is proportioned to the *volume* of cylinder behind the pixel.

Once an estimate of the matrix has been obtained, we obtain an estimate of ϕ by simply subtracting $\overline{U}_p^{(\text{est})}$ from the pth diagonal term of the matrix.

We shall test the method outlined above by computer simulation. We choose an ensemble consisting of just two "patients." Each element of the ensemble is a cylinder filled with isotope (see FIGURE 7). The height and the diameter of the cylinders are equal, so the projection of each cylinder onto the camera face is a square. The two "patients" differ in that the axis of the cylinder is vertical in one and horizontal in the other. This does not alter the square projection, but it does make a difference in the nuclear scan because each pixel in the camera records an expected number of counts proportional to the *volume* that projects onto the pixel in question. In fact we have

$$U_p = \rho_0 v_p t \quad (103)$$

where U_p is (as before) the expected number of counts in pixel p, ρ_0 is the number of disintegrating events per unit volume per unit time, v_p is the volume that projects onto pixel p, and t is the exposure time.

Thus the pixels aligned with the axis of the cylinder should receive the greatest numbers of counts because the volumes projecting onto them are greatest.

Since the ensemble is known in this case, we can compute the optimal filter directly without observing the scans of a large number of patients. Thus, we defer to future work the question of how many patient scans are needed to obtain an adequate version of the optimal filter. For this reason, the results obtained with the filter are labeled "ideal filter" in FIGURE 9.

FIGURE 8 shows the raw (unfiltered) data at three different exposures for the vertical and horizontal cylinders. The results have been normalized by the exposure so that the mean level is the same in all three cases, but there is less noise at the higher exposures. The discrimination task (between vertical and horizontal) is difficult at all three exposures, but it is certainly possible at the highest exposure. FIGURE 9 shows the improvement obtained by filtering the data at the middle exposure (20 sec). The discrimination between the vertical and horizontal cylinder is far easier with the filtered images than with the raw data; moreover, the filtered images are very close to those labeled ideal data, which are simply plots of the *expected* number of counts U_p in each of the two orientations.

CONCLUSIONS

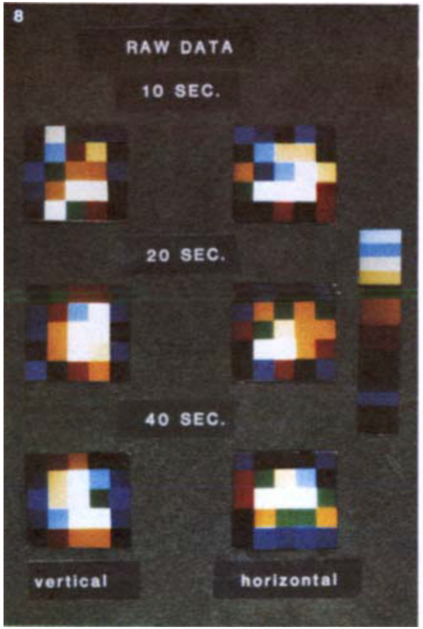
At low levels of light in vision, and at low dose of isotope in nuclear medicine, resolution is limited by the photon noise. This statement applies both to temporal resolution and also to spatial resolution. One strategy for dealing with the photon noise is to use a filter that attempts to reconstruct the image by smoothing each photon event and then summing the results. We have shown in this paper how the optimal smoothing can be obtained. This optimal filter always has the feature that it is faster and sharper when more photons are available.

An important and perhaps unexpected feature of the optimal filter is that its design is dependent on certain statistical information concerning the ensemble of scenes that the filter is designed to process. In the case of nuclear medicine, we have a concrete (albeit noisy) record of the ensemble in the form of similar nuclear medicine scans that have been obtained in the past. We have shown how such data can be used to extract the necessary statistical information concerning the ensemble. In the case of vision, it is much less clear what constitutes the ensemble of scenes. One possibility is that the retina of each species has adapted in the course of evolution to the *typical* scenes that are relevant for that species. Another possibility is that the retina functions as an adaptive filter, adjusting its transfer function to the statistical properties of the *current* retinal image.

We believe that the theory developed in this paper will be useful in explaining the changes in the behavior of the retina that occur as the ambient light level varies. We also believe that the class of optimal filters described in this paper will provide a practical method for processing nuclear medicine images. In the latter case we should be able to improve the image quality at any given dose or alternatively achieve the same quality with a lower dose of isotope.

SUMMARY

This paper is concerned with the construction of optimal filters for smoothing the photon noise that arises under low-light conditions in vision and under low-dose or



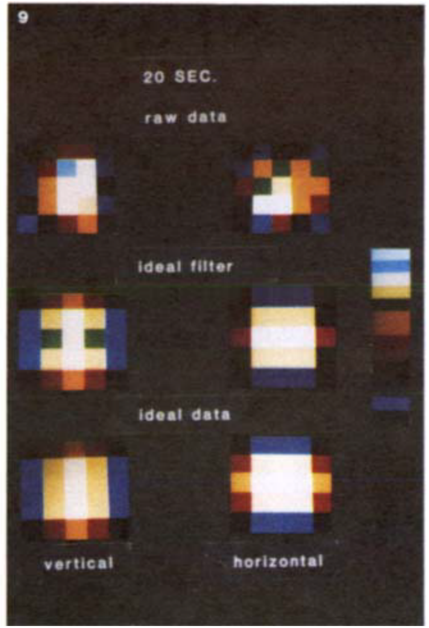


FIGURE 8. Simulated nuclear medicine scans of the two cylinders in FIGURE 7 at three different exposures. The numbers of counts have been normalized by dividing by the exposure time. As the exposure increases there is less noise, and it becomes easier to tell which cylinder is vertical and which is horizontal.

FIGURE 9. Performance of the optimal filter applied to simulated nuclear medicine scans. Top (raw data): unfiltered scans of the two cylinders at 20-sec exposure. Middle (ideal filter): the filtered scans. Note that the discrimination between vertical (left) and horizontal (right) is far easier with the filtered images. Bottom (ideal data): the *expected* number of counts for each cylinder. Note that the filtered scans are very similar to the ideal data.

short-exposure conditions in nuclear medicine. In the case of vision, the paper explores the hypothesis that the retina functions as an optimal filter. The consequences of this hypothesis for light adaptation are studied. In the case of nuclear medicine, a method for constructing optimal filters is introduced, and the method is tested by computer simulation.

ACKNOWLEDGMENTS

We gratefully acknowledge the helpful advice of R. M. Shapley and M. D. Blaufox.

REFERENCES

1. ROSE, A. 1948. The sensitivity performance of the human eye on an absolute scale. J. Opt. Soc. Am. **38:** 196–208.
2. SHAPLEY, R. & C. ENROTH-CUGELL. 1984. Visual adaptation and retinal gain controls. *In* Progress in Retinal Research. N. Osborne & G. Chader, Eds. Pergamon Press. London, England.
3. WIENER, N. 1949. Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications. MIT Press. Cambridge, Mass.
4. LEE, Y. W. 1960. Statistical Theory of Communication. John Wiley. New York, N.Y.
5. MORSE, P. M. & H. FESHBACH. 1953. Methods of Theoretical Physics, Vol. I, chapt. 8. McGraw-Hill. New York, N.Y.
6. TRANCHINA, D., J. GORDON & R. M. SHAPLEY. 1984. Retinal light adaptation—evidence for a feedback mechanism. Nature (London) **310:** 314–316.

DISCUSSION

DR. M. GREEN (*City College of CUNY, New York, N.Y.*): What is the definition of the diffusion constant in your transform in the gain loop for the retina?

DR. C. PESKIN: I think that you are asking how we pick the diffusion constant for the retina. It would come from the diffusion coefficient of the process, that is, *that* retina was designed for looking at *this* particular collection of scenes. Part of the definition of that process was that the fireflies (the spots of light) have a certain random walk with a certain diffusion coefficient. So the two would be matched to each other.

GREEN: In other words, it is not purely a retinal property. It is partly a property of the ensemble of scenes.

PESKIN: Yes. The fundamental point that I want to make is that an optimal filter is a function of the class of scenes at which you intend it to look. It obviously is not a function of the particular scene. If you knew the particular scene, you would not bother to make a filter, for you would know the answer already. But the important concept here is that one has a certain class of scenes that are related in some way and one can design a filter for that whole class to pick out the individual one as best possible.

DR. F. STRAND (*New York University, New York, N.Y.*): I have a more general question: What is happening in terms of retinal sensitivity to photons when there is a reversal of background in an optical illusion such as one in which the background first appears to be white and then black?

PESKIN: I am sorry, I don't know.