MASSACHUSETTS INSTITUTE OF TECHNOLOGY

A.I. LABORATORY

Artificial Intelligence
Memo. No. 295

October 1973

ON LIGHTNESS

Berthold K. P. Horn

The intensity at a point in an image is the product of the reflectance at the corresponding object point and the intensity of illumination at that point. We are able to perceive lightness, a quantity closely correlated with reflectance. How then do we eliminate the component due to illumination from the image on our retina? The two components of image intensity differ in their spatial distribution. A method is presented here which takes advantage of this to compute lightness from image intensity in a layered, parallel fashion.

The method is developed for a restricted class of images first used by Land in presenting his retinex theory of color. In this theory the problem of color perception is reduced to one of judging black and white lightness on three images taken in different parts of the visual spectrum. The method described here fills the need for a lightness judging process.

The theory has implications for potential special purpose hardware in image sensing devices. It should also be of interest to cognitive psychologists since it can explain certain effects observed in the human visual system as well as predict new ones. Further, the theory provides neuro-physiologists with suggestions about the function of certain structures in the primate retina.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-70-A-0362-0005.

Acknowlegments

I wish to thank Dr. David Marr for a number of most interesting discussions and for encouraging me to commit this method to paper. He first observed that the primate retina appears to have just the right detailed structure to implement the functions here described.

LIGHTNESS: Definition

The relative degree to which an object reflects light.

The Random House Dictionary

The attribute of object colors by which the object appears to reflect or transmit more or less of the incident light.

Webster's Seventh New Collegiate Dictionary

Preview

Part 1 is a review of the relevant information relating to color vision and lightness. This includes a discussion of the Land retinex model in a form suitable for the developments of the next part.

In part 2, Land's one-dimensional operation will be extended to two-dimensional images. The method depends on a layered, parallel computation suggestive of both biological and artificial implementations.

In part 3 some of the implications are explored and the information on the new image processing technique is summarized.

1. Review.

1.1 Theories of Color Perception.

There has always been great interest in how we perceive colors and numerous explanations have been forwarded (Newton 1704, Goethe 1810, Young 1820, Maxwell 1856, Helmholtz 1867, Hering 1875). The human perceptual apparatus is remarkably succesful in coping with large variations in the illumination. The colors we perceive are closely correlated with the surface colours of the objects viewed, despite large temporal and spatial differences in color and intensity of the incident light. This is surprising since we cannot sense reflectance directly.

The light intensity at a point in the image is the product of the reflectance at the corresponding object point and the intensity of illumination at that point - aside from a constant factor that depends on the optical arrangment. There must then be some difference between these two components of image intensity which allows us to discount the effect of one. The two components differ in their spatial distribution. Incident light intensity will usually vary smoothly, with no discontinuities, while reflectance will have sharp discontinuities at edges where objects adjoin. The reflectance being relatively constant between such edges.

1.1.1 Tri-Stimulus Theory.

Some facts about how we see color are fairly well established. It appears that we have three kinds of sensors operating in bright illumination, with peak sensitivities in different parts of the visible spectrum. This is why it takes exactly three colors in additive mixture to match an unknown color. While it is a bit tricky to measure the sensitivity curves of the three sensors directly, a linear transform of these curves has been known accurately for some time (Brindley 1960). These curves, called the standard observer curves, are sufficient to allow one to predict color matches made by subjects with normal colour vision (Hardy 1936).

The simplest theory of color perception then amounts to locally comparing the outputs of three such sensors and assigning colour on this basis (Young 1820, Helmholtz 1867). This however totally fails to explain the observed color contancy. Perceived color does not depend directly on the relative amounts of light measured by the three sensors (Land 1959, Lettvin 1967).

1.1.2 Color Conversion.

A number of attempts have been made to patch up this theory under the rubrics of "discounting of the illuminant", "contrast effect adjustment" and "adaptation". The more complicated theories are based on models with large numbers of parameters which are adjusted according

to empirical data (Helson 1938 & 1940, Judd 1940 & 1952, Richards 1971). These theories are at least partially effective in predicting human color perception when applied to simple arrangements of stimuli similar to those used in determining the parameters.

The parameters depend strongly on the data and slight experimental variations will produce large fluctuations in them. This is a phenomena familiar to numerical analysts fitting curves to data when the number of parameters is large. These theories are lacking in parsimony and convincing physiological counter-parts. Lettvin has demonstrated the hopelessness of trying to find fixed transformations from locally compared output of sensors to perceived color (Lettvin 1967).

1.2 Land's Retinex Theory.

Another theory of color perception is embodied in Land's retinex model (Land 1959, 1964 & 1971). Land proposes that the three sets of sensors are not connected locally, but instead are treated as if they represent points on three separate images. Processing is performed on each such image separately to remove the component of intensity due to illumination gradient. Such processing is not merely an added frill but is indispensible to color perception in the face of the variability of the illumination.

1.2.1 Lightness Judging.

In essence a judge of lightness processes each image. Lightness is the perceptual quantity closely correlated with surface reflectance. Only after this process can the three images be compared to reliably determine colors locally. It remains to mechanize this process.

It would appeal to intuition if this process could be carried out in a parallel fashion that does not depend on previous knowledge of the scene viewed. This is because colors are so immediate, and seldom depend on one's interpretation of the scene. Colors will be seen even when the picture makes no sense in terms of previous experience. Also, color is seen at every point in an image.

1.2.2 Mini-world of Mondrians.

In developing and explaining his theory Land needed to postpone dealing with the full complexity of arbitrary scenes. He selected a particular class of objects as inputs, modelled after the paintings of the turn-of-the-century Dutch artist Pieter Cornelis Mondrian. These scenes are flat areas divided into sub-regions of uniform matte color. Problems such as those occasioned by shadows and specular reflection are avoided in this way. One also avoids shading; that is, the variation in reflectance with the orientation of the surface in respect to the sensor and the light-source. For Mondrians, lightness is considered to be a function of reflectance.

Mondrians are usually made of polygonal regions with straight sides - for the development here however the edges may be curved. In the world of Mondrians one finds that the reflectance has sharp discontinuities wherever regions meet, being constant inside each region. The illumination, on the other hand, varies smoothly over the image.

1.3 Why Study the One-dimensional Case?

Images are two-dimensional and usually sampled at discrete points. For historic reasons and intuitive simplicity the results will first be developed in one dimension, that is with functions of one variable. Similarly, continous functions will be used at first since they allow a cleaner separation of the two components of image intensity and illustrate more clearly the concepts involved.

Use will be made of analogies between the one-dimensional and two-dimensional cases as well as the continous and discrete ones. The final process discussed for processing image intensities is two-dimensional and discrete. A number of physical implementations for this scheme are suggested.

The process will be looked at from a number of points of view: partial differential equations, linear systems, fourier transforms and convolutions, difference equations, iterative solutions, feed-back schemes and physical models.

1.3.1 Notation.

The following notation will be used:

s' Intensity of incident illumination at a point on the object.

r' Reflectance at a point of the object.

p' Intensity at an image point. Product of s' and r'.

s, r, p: Logarithms of s', r' and p' respectively.

d Result of applying forward or differencing operator to p.

t Result of applying threshold operator to d.

I Result of applying inverse or summing operator to t.

D Simple derivative operator in one dimension.

T Continous threshold operator, discards finite part.

I Simple integration operator in one dimension.

L Laplacian operator - sum of second partial derivatives.

G Inverse of the Laplacian, convolution with $(1/2\pi)\log_{\epsilon}(1/r)$.

D*, T*, I*, L* and G*: Discrete analogues of D, T, I, L and G.

The output I, will not be called lightness since there is probably not yet a generally acceptable definition of this term. It is however intended to be monotonically related to lightness. Note that I is

related to the logarithm of reflectance, while the perceptual quantity is perhaps more closely related to the square-root of reflectance.

1.4 One-Dimensional Method - Continous Case.

Land invented a simple method for separating the image components in one dimension. First one takes logarithms to convert the product into a sum. This is followed by differentiation. The derivative will be the sum of the derivatives of the two components. The edges will produce sharp pulses of area proportional to the intensity steps between regions - while the spatial variation of illumination will produce only finite values everywhere. Now if one discards all finite values, one is left with the pulses and hence the derivative of lightness. Finally one undoes the differentiation by simple integration.

1.4.1 One-Dimensional Continous Method: Details.

We have the following: Let $r'(x)$ be the reflectance of the object at the point corresponding to the image point x. Let $s'(x)$ be the intensity at this object point. Let $p'(x)$ be their product, that is, the intensity recorded in the image at point x. Note that $s'(x)$ and $r'(x)$ are positive.

$$p'(x) = s'(x) * r'(x)$$

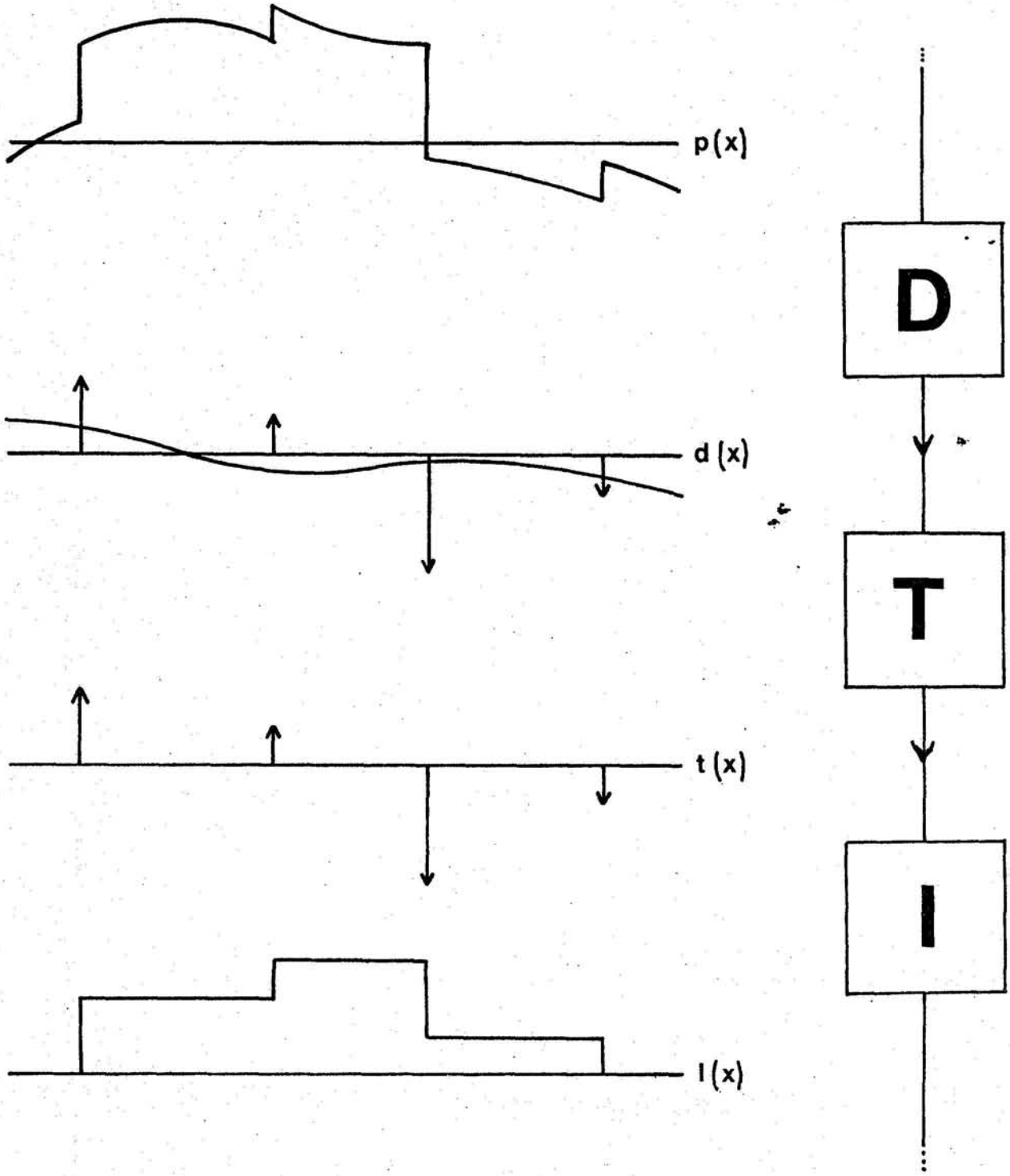

FIGURE 1: Processing steps in the one-dimensional continous case.

Now let p(x) be the logarithm of p'(x) and so on:

$$p(x) = s(x) + r(x)$$

Note that s(x) is continous and that r(x) has some finite discontinuities. Let D represent differentiation with respect to x.

$$d(x) = D(p(x)) = D(s(x)) + D(r(x))$$

Now, D(s(x)) will be finite everywhere, while D(r(x)) will be zero aside from a number of pulses - which carry all the information. Each pulse will correspond to an edge between regions and have area proportional to the intensity step. If now one "thresholds" and discards all finite parts, one gets:

$$t(x) = T(D(p(x))) = D(r(x))$$

To obtain r(x) one only has to invert the differentiation, that is, integrate. Let I represent integration with respect to x, then $(I)^{-1} = D$ and:

$$I(x) = I(T(D(p(x))) = r(x) + c$$

One can give a convolutional interpretation to the above, since differentiation corresponds to convolution with a pulse-pair, one

negative and one positive, each of unit area. Integration corresponds to convolution with the unit step function.

1.4.2 Normalization.

The result is not unique because of the constant introduced by the integration. The zero (spatial) frequency term has been lost in the differentiation, so cannot be reconstructed. This is related to the fact that one does not know the overall level of illumination and hence cannot tell whether an object appears dark because it is grey or because the level of illumination is low.

One can normalize the result if one assumes that there are no light sources in the field of view and no flourescent colors or specular reflections. This is certainly the case for the Mondrians. Perhaps the best way of normalising the result is to simply assume that the highest value of lightness corresponds to white, or total reflectance in the Lambertian sense. This normalization will lead one astray if the image does not contain a region corresponding to a white patch in the scene, but this is the best one can do. Other normalization techniques might involve adjusting weighted local averages, but this would then no longer amount to reconstruction of reflectance.

1.5 One-Dimensional Method - Discrete Case.

So far we have assumed that the image intensity was a continous function. In retinas found in animals or artificial ones constructed out of discrete components, images are only sampled at discrete points. So one has to find discrete analogues for the operations we have been using. Perhaps the simplest are first differences and summation as analogues of differentiation and integration respectively. This is not to say that other approximations could not be used equally well.

To use the new operators one goes through essentially the same process as before, except that now all values in the differenced image are finite. This has the effect of forcing one to choose a threshold for the thresholding function. Both components of image intensity produce finite values after the differencing operation. The component due to the edges in the reflectance is hopefully quite large compared to that due to illumination gradient. One has to find a level that will suppress the illumination gradient inside regions, while permitting the effects due to edges to remain.

1.5.1 One-Dimensional Discrete Method: Details.

Let r'_i be the reflectance of the object at the point corresponding to the image point i. Let s'_i be the incident light intensity at this object point. Let p'_i be their product, that is the intensity in the image at point i.

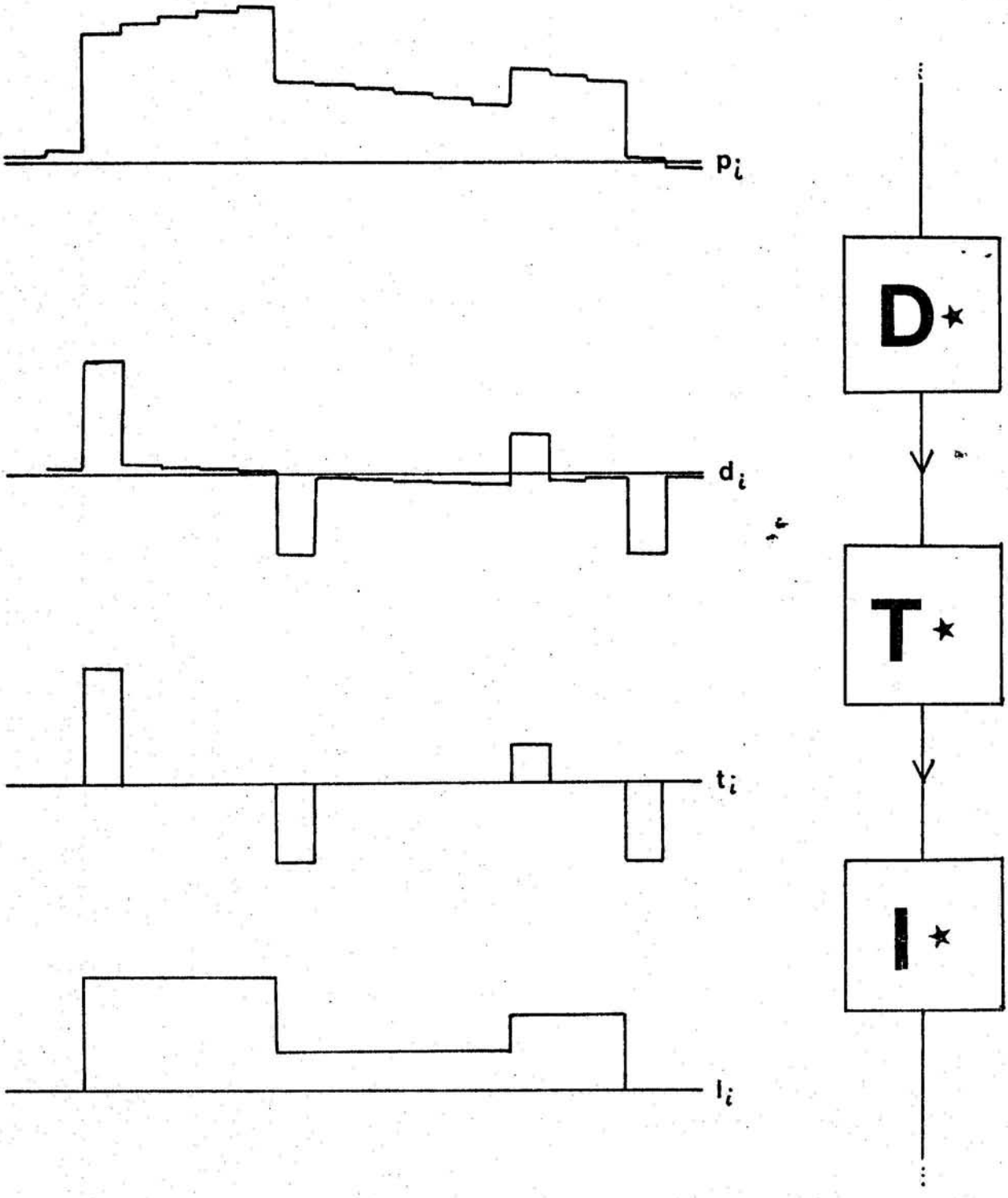

FIGURE 2: Processing steps in the one-dimensional discrete case.

$$p'_i = s'_i * r'_i$$

Now let p_i be the log of p'_i and so on. Let D* and I* be the operators corresponding to taking first differences and summation respectively. Note that $(I*)^{-1} = D*$.

$$P_i = s_i + r_i$$

$$d_i = p_{i+1} - p_i \qquad (d = D*(p))$$

$$t_i = d_i \text{ if } |d_i| < e, \text{ else } 0$$

$$l_i = \sum_{k=0}^{i} t_k \qquad (l = I*(t))$$

1.5.2 Selecting the Threshold.

What determines the threshold? It must be smaller than the smallest intensity step between regions. It must on the other hand be larger than values produced by first differencing the maximum illumination gradients. Real images are noisy and the threshold should be large enough to eliminate this noise inside regions.

The spacing of the sensor cells must also be taken into account.

As this spacing becomes smaller, the contribution due to illumination gradients decreases, while the component due to the edges remains constant. A limit is reached when the component due to illumination gradients falls below that due to noise or when the optical properties of the imaging system begin to have a deliterious effect. In all imaging systems an edge is spread over a finite distance due to diffraction and uncorrected abberations. The spacing of sensors should not be much smaller than this distance to avoid reducing the component due to edges in the differenced image.

Let u be the radius of the point-spread-function of the optical system and h the spacing of the sensor cells. Let g' be the smallest step in the logarithm of reflectance in the scene. Then define the effective minimum step as:

$$g = g' * \min(1, h/2u)$$

Let α be the largest slope due to illumination gradient and σ the root-mean-square noise-amplitude. The noise will exceed a value 3σ only .3% of the time. Choose the threshold e as follows:

$$e < g$$

$$e > \alpha * h$$

$$e > 3\sqrt{2}\ \sigma$$

1.5.3 Accuracy of the Reconstruction.

In the continous case one can exactly reconstruct the reflectance, aside from a constant. We are not so fortunate here, even if we select a threshold according to the above criteria. This is because the values at the edges contain small contributions due to illumination gradient and noise. A slight inaccuracy in the reconstruction will result. This error is minimized by making the sensor cell spacing very fine, optimally of a size commensurate with the optical resolution of the device. The effect of noise can also be minimized by integrating over time.

Note that the reconstruction is more accurate when there are few edges, since it is at the edges that the error effects appear. With many edges the illumination gradient begins to "show through".

1.5.4 Generalizations.

So far we have dealt with constant sensor spacing. Clearly as long as the same spacing is used for both the differencing and the summing, the cell spacing can be arbitrary and has little effect on the reconstruction since it does not enter into the equations.

Similarly we have chosen first differences as the discrete analogue for differentiation. We could have chosen some other weighted difference and developed a suitable inverse for it. This inverse of course would no longer be summation but can be readily obtained using

techniques developed for dealing with difference equations (Richtmeyer 1957, Garabedian 1964).

1.5.5 Physical Models of the One-Dimensional Discrete Process.

One can invent a number of physical models of the above operations. A simple resistive network will do for the summation process for example. Land has implemented a small circular "retina" with about 16 sensors. This model employs electronic components to perform the operations of taking logarithms, differencing, thresholding and summing.

Land has tried to extend his one-dimensional method to images, by covering the image with paths produced by a random walk procedure and applying methods like the above to each of these paths. While this produces results, it seems unsatisfactory from the point of view of suggesting possible neuro-physiological structures: neither does it lend itself to efficient implementation.

Methods depending on non-linear processing of the gradient along paths in the image fail to smoothly generalize to two dimensions, and cannot predict the appearance of images in which different paths result in different lightnesses.

2. Lightness in Two Dimensional Images.

2.1 Two-Dimensional Method - Continous Case.

We need to extend our ideas to two dimensions in order to deal with actual images. There are a number of ways of arriving at the process to be described here, we shall follow the simplest (Horn 1968). We need to find two-dimensional analogues to differentiation and integration. The first partial derivatives are directional and thus unsuitable since they will for example completely eliminate evidence of edges running in a direction parallel to their direction of differentiation. Exploring the partial derivatives and their linear combinations one finds that the Laplacian operator is the lowest order combination that is isotropic, or rotationally symmetric. The Laplacian operator is of course the sum of the second partial derivatives.

2.1.1 Applying the Laplacian to a Mondrian.

Before investigating the invertibility of this operator, let us see what happens when one applies it to the image of a Mondrian. Inside any region one will obtain a finite value due to the variation in illumination intensity. At each edge one will get a pulse pair, one positive and one negative. The area of each pulse will be equal to the intensity step.

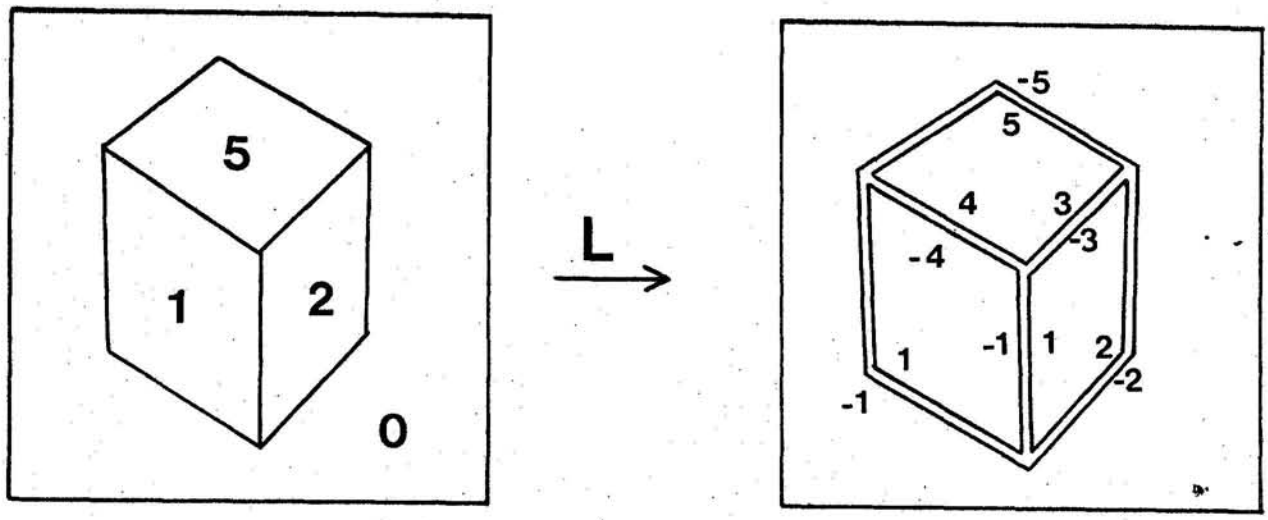

FIGURE 3: Applying the laplacian operator to the image of a Mondrian figure.

This can best be seen by considering the first derivative of a step, namely a single pulse. If this is differentiated again one obtains a doubled pulse as described. Since this pulse will extend along the edge, one may think of it as a pulse-wall. So each edge separating regions will produce a doubled pulse wall. It is clear that one can once again separate the component due to reflectance and illumination simply by discarding all finite parts.

2.1.2 Inverse of the Laplacian Operator.

To complete the task at hand one then has to find a process for undoing the effect of applying the Laplacian. Again there are a number of approaches to this problem, we will use the shortest (Horn 1968). In essence one has to solve for p(x,y) in a partial differential equation of the form:

$$L(p(x,y)) = d(x,y)$$

This is Poisson's equation and it is usually solved inside a bounded region using Green's function (Garabedien 1964):

$$p(x,y) = \iint G(\xi,\eta;x,y) * d(\xi,\eta)\, d\xi\, d\eta$$

The form of Green's function G, depends on the shape of the region boundary. Now if the retina is infinite all points are treated similarly and Green's function depends only on two parameters, $(\xi - x)$ and $(\eta - y)$. This positional independence implies that the above integral simply becomes a convolution. It can be shown that Green's function for this case is:

$$G(\xi, \eta ; x, y) = (1/2\pi) \log_e(1/r)$$

Where $r^2 = (\xi - x)^2 + (\eta - y)^2$

So $$p(x,y) = \int\int (1/2\pi) \log_e(1/r) * d(\xi, \eta) \, d\xi \, d\eta$$

Thus the inverse of the Laplacian operators is simply convolution with $(1/2\pi) \log_e(1/r)$. To be precise one has:

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) \int\int (1/2\pi) \log_e(1/r) * d(\xi, \eta) \, d\xi \, d\eta = d(x,y)$$

This is the two-dimensional analogue of:

$$\frac{d}{dx} \int_{-\infty}^{x} f(t)\, dt = f(x)$$

2.1.3 Why one can use the Convolutional Inverse.

If the retina is considered infinite one can express the inverse as a simple convolution. If the retina is finite on the other hand one has to use the more complicated Green's function formulation.

Now consider a scene on a uniform background whose image is totally contained on the retina. The result of applying the forward transform and thresholding will be zero in the area of the uniform background. The convolutional inverse will therefore receive no contribution from outside the retina. As a result one can use the convolutional form of the inverse provided the image of the scene is totally contained within the retina.

2.1.4 Normalization.

Once again one finds that the reconstructed reflectance is not unique. That is, any non-singular solution of $L(p(x,y)) = 0$ can be added to the input without affecting the result. On the infinite plane such solutions have the form $p(x,y) = (a*x + b*y + c)$. If the scene only occupies a finite region of space it can be further shown that the solution will be unique up to a constant and that one does not have to

worry about possible slopes. To be specific: the background around the scene will be constant in the reconstruction. So one has here exactly the same normalization problem as in the one-dimensional case. Assigning white to the region with highest numerical value in the reconstructed output appears to be a reasonable method.

2.1.5 Two-Dimensional Continous Method: Details.

Let $r'(x,y)$ be the reflectance of the object at the point corresponding to the image point (x,y). Let $s'(x,y)$ be the source intensity at that object point. Let $p'(x,y)$ be their product, that is the intensity at the image point (x,y). Note that $r'(x,y)$ and $s'(x,y)$ are positive.

$$p'(x,y) = s'(x,y) * r'(x,y)$$

Let $p(x,y)$ be the logarithm of $p'(x,y)$ and so on.

$$p(x,y) = s(x,y) + r(x,y)$$

Now assume that $s(x,y)$ and its first partial derivatives are continous - a reasonable assumption to make for the distribution of illumination on the object. Let L be the Laplacian operator.

$$d(x,y) = L(p(x,y)) = L(s(x,y)) + L(r(x,y))$$

Now $L(s(x,y))$ will be finite everywhere, while $L(r(x,y))$ will be zero except at each edge separating regions, where one will find a double pulse wall as described. Now discard all finite parts:

$$t(x,y) = T(L(p(x,y))) = L(r(x,y))$$

Let G be the operator corresponding to convolution by $(1/2\pi)\ \log_e(1/r)$. Note that $(G)^{-1} = L$.

$$I(x,y) = G(T(L(p(x,y))) = r(x,y) + c$$

2.2 Two-Dimensional Method - Discrete Case.

Once again we turn from a continous image to one sampled at discrete points. First we will have to decide on a tesselation of the image plane.

2.2.1 Tesselation of the Image Plane.

For regular tesselations the choice is between triangular, square and hexagonal unit cells. In much past work on image processing, square tesselations have been used for the obvious reasons. This particular tesselation of the image has a number of disadvantages. Each cell has

two kinds of neighbors, four adjoining the sides, four on the corners. This results in a number of asymmetries. It makes it difficult to find convenient difference schemes approximating the Laplacian operator with low error term.

Triangular unit cells are even worse in that they have three kinds of neighbors, compounded these drawbacks. Note also that near-circular objects pack tightest in a pattern with hexagonal cells. For these reasons we will use a hexagonal unit cell. It should be kept in mind however that it is easy to develop equivalent results using different tesselations.

2.2.2 Discrete Analogue of the Laplacian.

Having decided on the tesselation we need now to find a discrete analogue of the Laplacian operator. Convolution with a central positive value and a rotationally symmetric negative surround of equal weight is one possibility. Aside from a negative scale factor, this will approach application of the Laplacian in the limit as the cell size tends to zero.

If one were to use complicated surrounds, the trade-offs between accuracy and resolution would suggest using a negative surround that decreases rapidly outward. For the sake of simplicity we will choose convolution with a central cell of weight 1, surrounded by six cells of weight -1/6. This function is convenient, symmetric and has a small error term. It is equal to - ($h/4\ L + h/64\ L^2$) plus sixth and higher

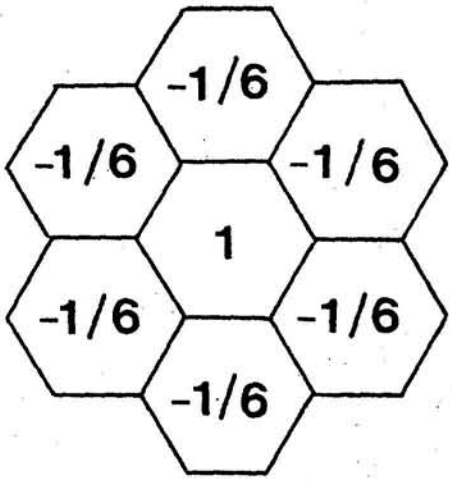

FIGURE 4a: A discrete analogue of the laplacian operator.

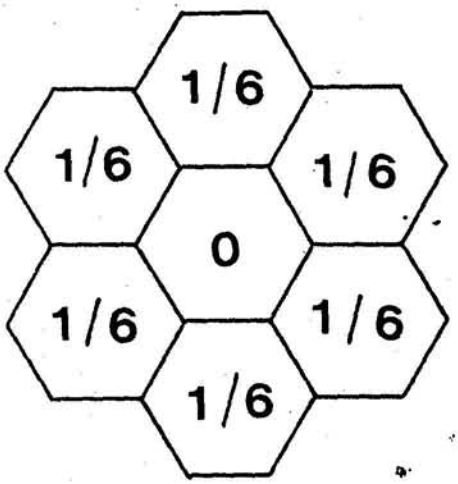

Figure 4b: Delta function minus this discrete analogue.

order derivatives (Richtmeyer 1957). It should again be pointed out that similar results can be developed for different functions.

2.2.3 Inverse of the Discrete Operator.

The forward differencing operator has the form:

$$d_{ij} = p_{ij} - \sum w_{k-i,l-j} * p_{kl}$$

Where p_{ij} is the logarithm of image intensity, w_{ij} are weights, which in our case are 1/6, and the sum is taken over the six immediate neighbors.

We now have to determine the inverse operation that recovers p_{ij} from d_{ij}. One approach is to try and solve the difference equation of the form:

$$p_{ij} - \sum w_{k-i,l-j} p_{kl} = d_{ij}$$

Or in matrix form: $W \underset{\sim}{p} = \underset{\sim}{d}$. Note that W is sparse, having 1's on the diagonal and -1/6's scattered around. For a finite retina with n sensor cells one has to introduce boundary conditions to ensure that one has as many equations as there are unknowns. One then simply inverts the

matrix W and gets: $\underset{\sim}{p} = W^{-1} \underset{\sim}{d}$.

This is entirely analogous to the solution in the continous case for a finite retina. W^{-1} corresponds to the Green's function. Much as Green's function has a large "support", that is, is non-zero over a large area, so W^{-1} is not sparse. This implies that a lot of computation is needed to perform the inverse operation.

2.2.4 Computational Effort and Simplification.

Solving the difference equations for a given image by simple Gauss-Jordan elimination requires of the order of $n^3/2$ arithmetic operations. Another approach is to invert W once and for all for a given retina. For each image then one needs only about n^2 arithmetic operations. Note that the other operations, such as forward differencing, require only about 6*n arithmetic operations.

What in effect is happening is that each point in the output depends on each point in the differenced image. Both have n points, so n^2 operations are involved. Not only does one have to do a lot of computation, but must also store up the matrix W^{-1} of size n^2. This is quite prohibitive for even a small retina.

This latter problem can be avoided if one remembers the simplification attendant to the use of an infinite retina in the continous case. There we found that the integral with Green's function simplified into a convolution. Similarly, if one assumes an infinite retina here, one finds that W and its inverse become very regular. The

rows in W are then all the same and the same is true of W^{-1}. Each value in the output then depends in the same way on the neighboring points in the differenced image. One need only store up the dependence of one point on its neighbors for this simple convolutional operation.

The only remaining difficulty is that W is now infinite and one can no longer invert it numerically - one has to find an analytical expression for the inverse. I have not been able to find this inverse exactly. A good first approximation is $\log_6(r_o/r)$ - except for $r = 0$, when one uses $1 + \log_6(r_o)$. Here r is the distance from the origin and r_o is arbitrary. The remainder left over when one applies the forward difference scheme to this approximation lies between $\log_6(1 + r^{-6})$ and $\log_6(1 - r^{-6})$. This error term is of the order of r^{-6}.

In practice one does not have an infinite retina, but as has been explained for the continous case one can use the convolutional method described above for a finite retina, provided that the image of the scene is wholly contained inside the boundaries of the retina. It is possible to find an accurate inverse of this kind valid for a limited retinal size by numerical means.

2.2.5 Two-Dimensional Discrete Method: Details.

Let r'_{ij} be the reflectance at the object point corresponding to the image point (i, j). Let s'_{ij} be the intensity of the incident light at this object point. Let p'_{ij} be the intensity in the image at point (i, j).

$$p'_{ij} = s'_{ij} * r'_{ij}$$

Let p_{ij} be the logarithm of p'_{ij} and so on. Let L* be the operator that corresponds to convolution with the analogue of the Laplacian. Let G* be its inverse.

$$p_{ij} = s_{ij} + r_{ij}$$

$$d_{ij} = p_{ij} - \sum w_{k-i,l-j} * p_{kl} \qquad (d = L*(p))$$

The weigths w_{ij} are 1/6 in this case, and the sum is taken over the six immediate neighbors.

$$t_{ij} = d_{ij} \quad \text{if } |d_{ij}| > e, \text{ else } 0$$

$$l_{ij} = \sum v_{k-i,l-j} t_{kl} \qquad (l = G*(t))$$

Here the sum extends over the whole retina and v_{ij} is the convolutional inverse found numerically as explained above.

2.2.6 Simplicity of the Inverse.

The forward transform, involving only a simple subtraction of immediate neighbors, is clearly a rapid, local operation. The inverse on the other hand is global, since each point in the output depends on each point in the differenced image. Computationally this makes the inverse slow. The inverse is simple in one sense however: The difference equations being solved by the inverse have the same form as the equations used for the forward transform and are thus local. The problem is that the output here feeds back into the system and effects can propagate across the retina. The apparent global nature of the inverse is thus of a rather special kind and, as we will see later, gives rise to very simple implementations involving only local connections.

2.2.7 Iterative Methods of Solution.

There are of course other methods for solving large sets of equations. The fact that W is sparse and has large diagonal elements, suggests trying something like Gauss-Seidel iteration. Each iteration takes about 6*n arithmetic operations. For effects to propagate across the retina one requires at least $\sqrt{(4*n^2 - 1)/3}$ iterations. This is because a hexagonal retina of width m has $(3*m^2 + 1)/4$ cells. The above suggests that one might be able to get away with less than n^2 arithmetic operations. In practice it is found that effects propagate very slowly

and many more iterations are needed to stabilize the solution. One does not have to store W, since it is easily generated as one goes along.

Iterative schemes correspond to adding a time-derivative to the Poisson equation and so turning it into a heat-equation. As one continues to iterate the steady-state solution is approached. This intuitive model gives some insight into how the process will converge.

2.2.8 Convergence of Iterative and Feed-back Schemes.

It is not immediately clear that iterative schemes of solving the difference equations will converge. If they do, they will converge to the correct solution. Let δ be the delta function, that is, one at the origin, zero elsewhere. It can be shown that if the forward convolutional operator is W, the convergence of iterative schemes depends on the behaviour of the error term, $(\delta - W)^n$, as n becomes large. Raising a convolutional operator to an integer power is intended to signify convolution with itself.

In our case, W is one at the origin, with six values of -1/6 around it. So $(\delta - W)$ will be zero at the origin with six values of 1/6 around it. Now while $(\delta - W)^n$ will always have a total area of one, it does spread out and its value tends to zero at every point as n tends to infinity. So this iterative scheme converges: similar results could be derived for other negative surrounds.

2.2.9 Setting the Threshold.

In the discrete case a finite threshold must be selected. As before, let g' be the smallest step in the logarithm of reflectance in the scene, h the sensor spacing and u the radius of the point-spread function of the optical system. Then we define the effective minimum step as:

$$g = g' * \min(1, h/2u)$$

There are some minor differences in what follows depending on whether one considers the sensor outputs to be intensity samples at cell-centers or averages over the cell area. The smallest output due to an edge will be about g/6. This is produced when the edge is oriented to cover just one cell of the neighborhood of six. Let β be the maximum of the intensity gradient - that is the Laplacian of intensity in this case. Choose the threshold e as follows:

$$e < g/6$$

$$e > \beta * h^2$$

$$e > 3\sqrt{7/6}\ \sigma$$

2.2.10 Some Notes on This Method.

Notice that an illumination gradient that varies as some power of distance across the image becomes a linear slope after taking logarithms and thus produces no component after the differencing operations. Such simple gradients are suppressed even without the thresholding operation.

In practice the parameters used in choosing the threshold may not be known or may be variable. In this case one can look at a histogram of the differenced image. It will contain values both positive and negative corresponding to edges and also a large number of values clustered around zero due to illumination gradients, noise and so on. The threshold can be conveniently choosen to contain this central blob.

Noise and illumination gradients have an effect similar to that in the one-dimensional case. With finite cell spacing one cannot precisely separate the two components of the image intensity and at each edge the information will be corrupted slightly by noise and illumination gradient. As the density of edges per cell area goes up the effect of this becomes more apparent. In highly textured scenes the illumination gradient is hard to eliminate.

Once again one has to decide on a normalization scheme. The best method probably is to let the highest numerical value in the reconstructed output correspond to white.

2.2.11 Dynamic Range Reduction.

Applying the retinex operation to an image considerable reduces the range of values. This is because the output, being related to reflectance, will only have a range of one to two orders of magnitude, while the input will also have illumination gradients. This will make such processing useful for picture recording and transmission (Horn 1968).

2.2.12 A Frequency Domain Interpretation.

It may be of interest to look at this method from yet another point of view. What one does is to exentuate the high-frequency components, threshold and then attenuate the high-frequency components. To see this, consider first the forward operation. The fourier transform of the convolutional operator corresponding to differentiation is $i\omega$. Similarly the two-dimensional fourier transform of the convolutional operator corresponding to the Laplacian is $-\rho^2$. Here ρ is the radius in a polar coordinate system of the two-dimensional frequency space. In either case one is multiplying the fourier transform by some function that increases with frequency. Now consider the reverse operation. The fourier transform of the convolutional operation corresponding to integration is $1/i\omega$. Similarly the fourier transform of $(1/2\pi)\ \log_e(1/r)$ is $-1/\rho^2$. So in the inverse step one undoes exactly the emphasis given to high frequency components in the forward operation.

In both the one-dimensional and the two-dimensional case one loses the zero frequency component. This is why the result has to be normalized.

2.3 Physical Models.

There are numerous continous physical models to illustrate the inverse transformation. Anything that satisfies Poisson's equation will do. Such physical models help one visualize what the inverse of a given function might be. Examples in two dimensions are: perfect fluid-flow, steady diffusion, steady heat-flow, deformation of an elastic membrane, electro-statics and current flow in a resistive sheet. In the last model for example, the input is the distribution of current flowing into the resistive sheet normal to its surface, the output is the distribution of electrical potential over the surface.

In addition to helping one visualize solutions, these continous models also suggest discrete models. These can be arrived at simply by cutting up the two-dimensional space in a pattern corresponding to the interconnection of neighboring cells. That is, the remaining parts form a pattern dual to that of the sensor cell pattern. We will discuss only one such discrete model.

2.3.1 A Discrete Physical Model.

Consider the resistive sheet described, cut up in the dual pattern of the hexagonal unit cell pattern. What will be left is an interconnection of resistors in a triangular pattern. The inputs to this system will be currents injected at the nodes, the potential at the nodes being the output. This then provides a very simple analog implementation of the tedious inverse computation.

It is perhaps at first surprising to see that each cell is not connected to every other in a direct fashion. One would expect this from the form of the computational inverse. Each cell in the output does of course have a connection via the other cells to each of the inputs. Paths are shared however in a way that makes the result both simple and planar.

Consider for the moment just one node. The potential at the node is the average of the potential of the six nodes connected to it plus the current injected times R/6, where R is the resistance of each resistor. The economy of connection is due to the fact that the outputs of this system are fed back into it. It also illustrates that this model locally solves exactly the same difference equation as that used in the forward transform, only now in reverse.

This immediately suggests an important property of this model: By simply changing the interconnections one can make an inverse for other forward transforms. Simplest of all are other image plane tesselations, both regular and irregular. One simply connects the resistors in the

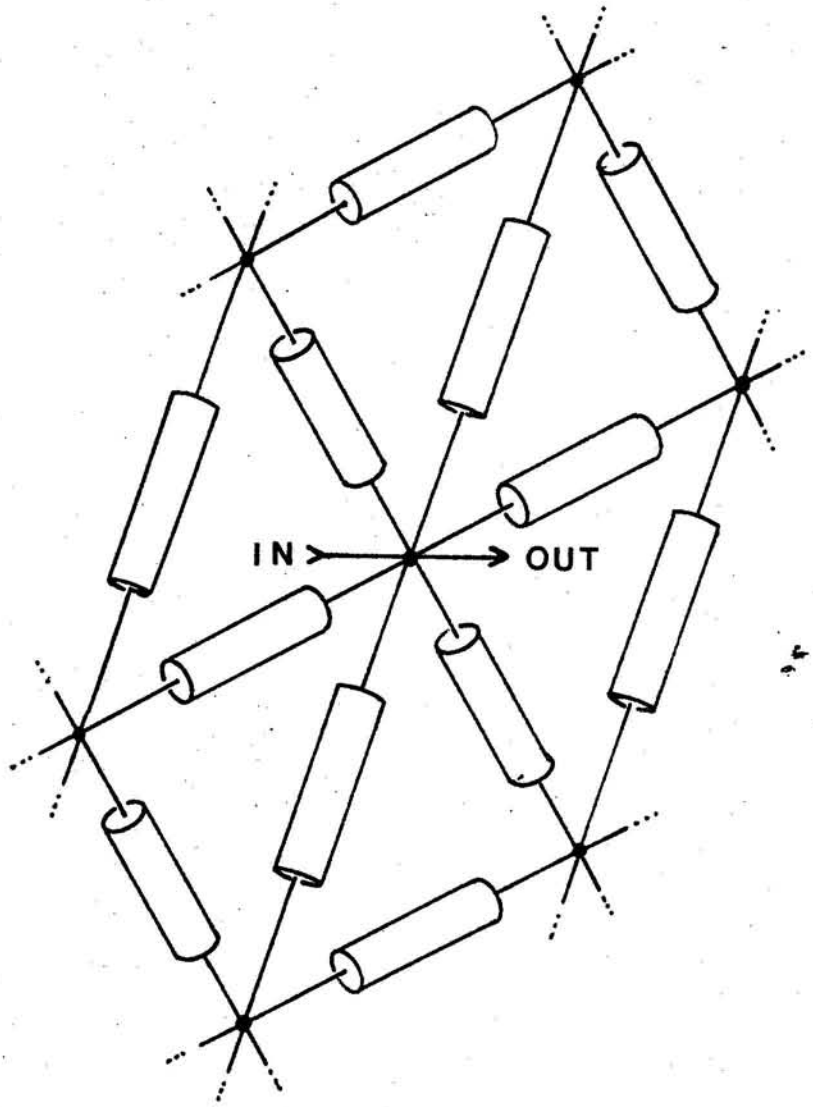

FIGURE 5: Resistive model of the inverse computation.
The inputs are the currents injected at the nodes.
The outputs are the potentials at the nodes.

same pattern as are the cells in the input.

More complicated weighted surrounds can be handled by using resistors with resistances inversely proportional to the weights. The network of resistors will then no longer be planar.

2.3.2 A Feed-back Scheme for the Inverse.

Both the comment about outputs feeding back into the resistive model and the earlier notes about iterative schemes suggest yet another interesting model for the inverse using linear summing devices. Operational amplifiers can serve this purpose. One simply connects the summing element so that they solve the difference equation implied by the forward transform. Once again it is clear that such a scheme can be generalized to arbitrary tesselations and weighted negative surrounds simply by changing the interconnections and attenuations on each input. Some questions of stability arise with esoteric interconnections. For the simple ones stability is assured.

A little thought will show that the resistive model described earlier is in fact a more economical implementation of just this scheme with the difference that there the inputs are currents, while here they are potentials.

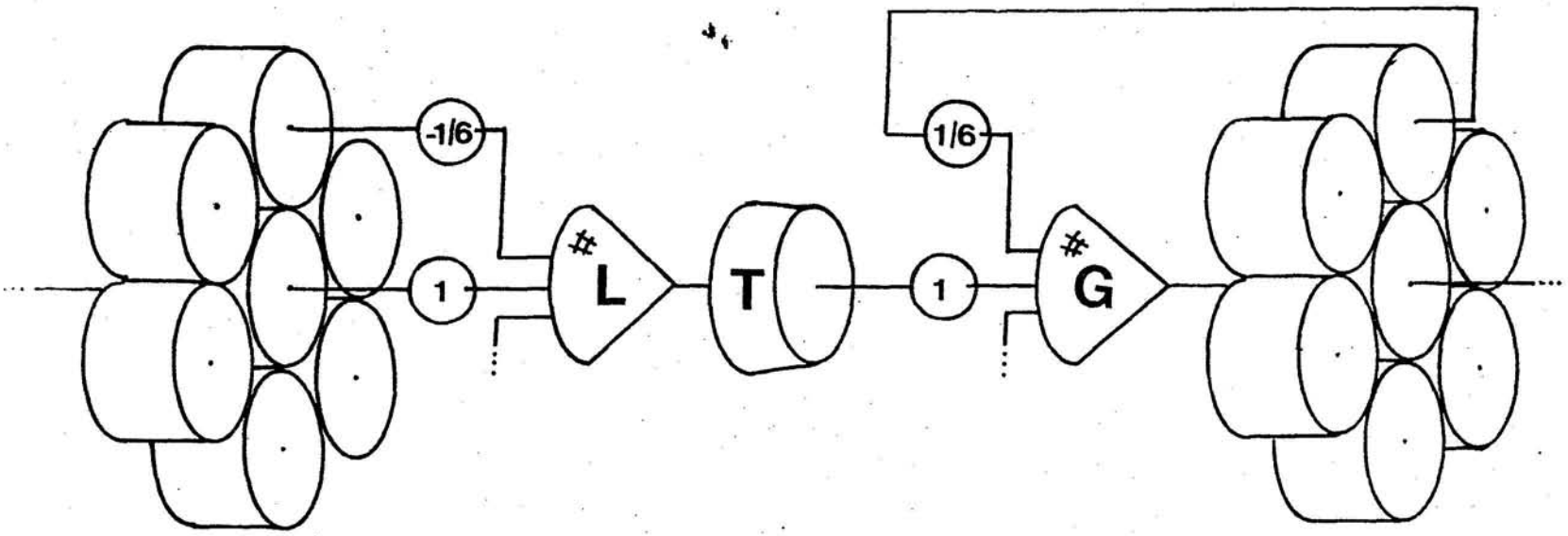

FIGURE 6: The use of summing elements and feed-back in the implementation of both the forward and the inverse transform.

2.4 Limitations of the Simple Scheme Presented.

The method presented here will not correctly calculate reflectance if used unmodified on general scenes. It may however calculate lightness fairly well. As the method stands now for example, a sharp shadow edge will not be distinguished from a real edge in the scene and the two regions so formed will produce different outputs, while their reflectances are the same. It may be that this is reasonable nevertheless, since we perceive a difference in apparent lightness.

Smooth gradations of reflectance on a surface due either to shading or variations in surface reflectance will be eliminated by the thresholding operations except as far as they affect the intensity at the borders of the region. This may imply that we need additional channels in our visual system to complement the ones carrying the retinexed information since we do utilize shading as a depth-cue.

The simple normalization scheme described will also be sensitive to specular reflections, flourescent paints and light-sources in the field of view. Large depth-discontinuities present another problem. One cannot assume that the illumination is equal on both sides of the obscuring edge. In this case the illuminating component does not vary smoothly over the retina, having instead some sharp edges.

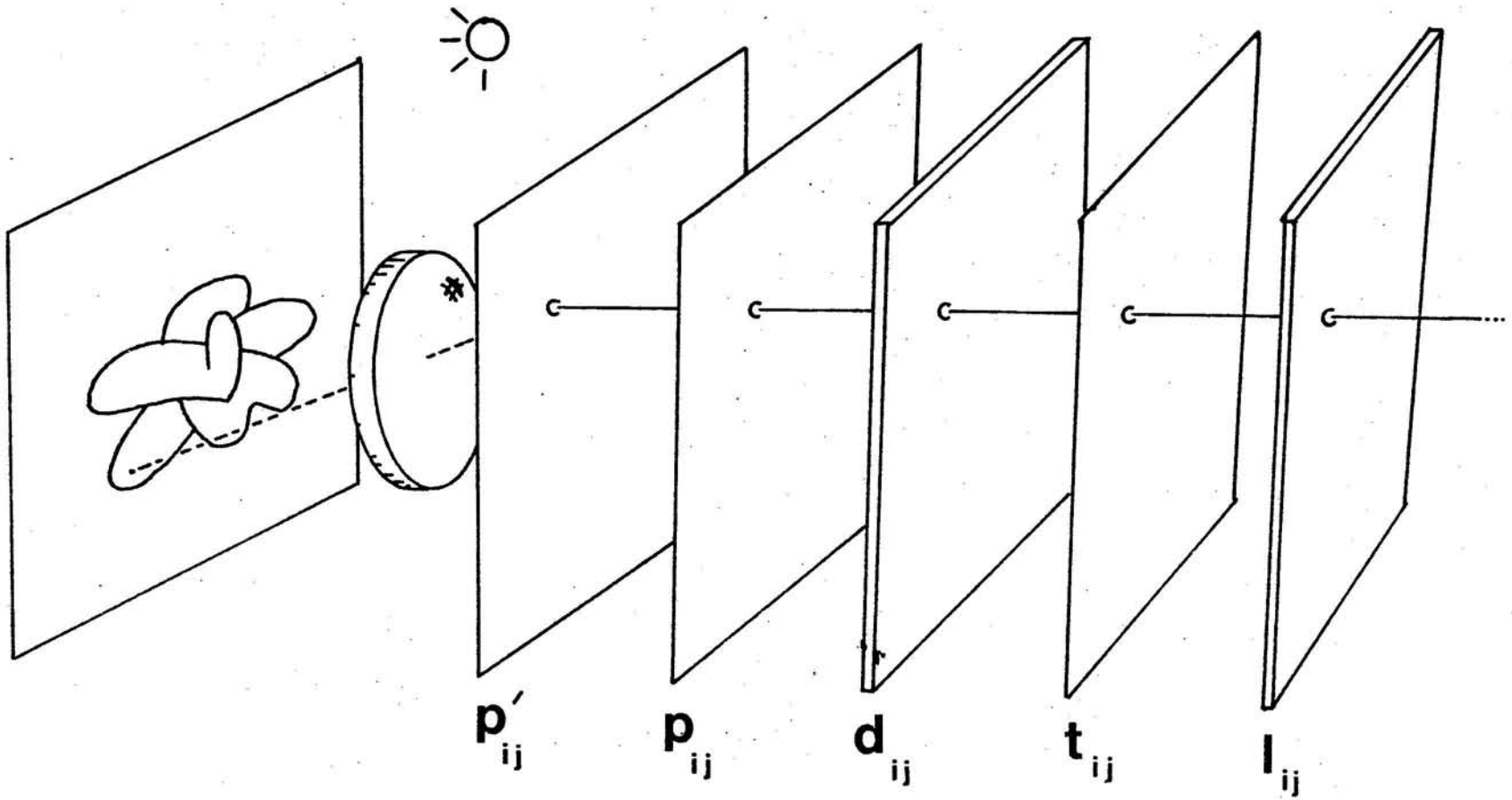

FIGURE 7: Illustration of the parallel layers of operations which produce the two-dimensional retinex operation. Only two of the operations involve local interactions between neighboring cells.

2.5 Computer Simulation of the Discrete Method.

A computer program was used to simulate the retinex process described on a small retina with both artificial and real images seen through an image dissector camera. The hexagonal unit cell is used in this program and the retina itself is also hexagonal. The retina contains 1027 cells in a pattern 36 cells across. This is a compromise dictated by the need to limit the number of arithmetic operations in the inverse transform. In this case one needs about a million and this takes about a minute of central processor time on our PDP-10.

Both the artificial and the real Mondrians consist of regions bounded by curved outlines to emphasize that this method does not require straight-line edges or boundary extraction and description. Various distributions of incident illumination can be selected for the artificial scenes. In each case the processing satisfactorily removes the gradient.

For the real scenes it is hard to produce really large illumination gradients by positioning the light-sources. The reconstruction does eliminate the gradient well, but often minor flaws will appear in the output due to noise in the input and a number of problems with this kind of input device such as a very considerable scatter. It is not easy to predict what effects such imaging device defects will have.

The output is displayed on a DEC 340 display which has a mere eight grey-levels. It would be interesting to experiment with larger retinas and better image input- and output-devices.

2.5.1 Form of Inverse used in the Computer Simulation.

The convolutional form of the inverse was used for speed and low storage requirement. This neccesitated solving the difference equations once, given a pulse as input. The symmetry of the hexagonal pattern allows one to identify symmetrically placed cells and only 324 unknowns needed to be found for a convolutional inverse sufficient for the size of retina described. As mentioned before, this function is closely approximated by $\log_6(r_o/r)$ for large r. This can be used to establish boundary conditions.

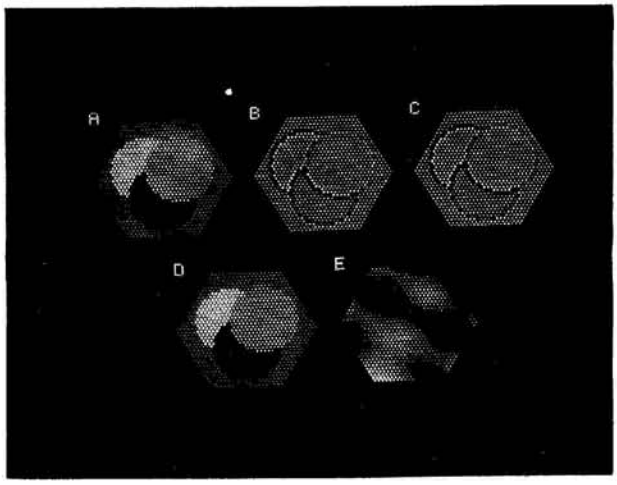

FIGURE 8: The method applied to an artificial image.

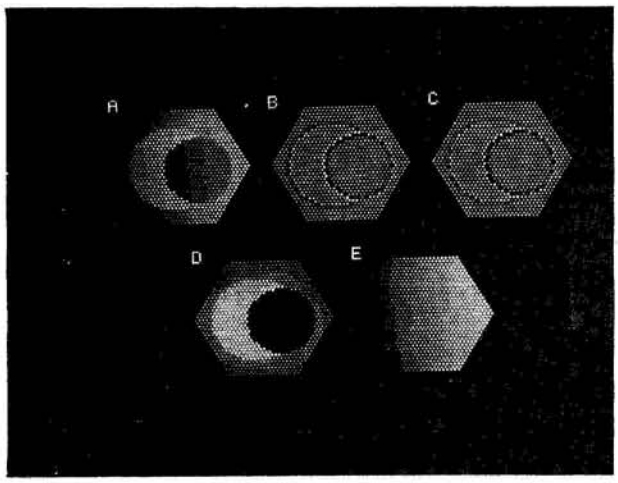

FIGURE 9: The method applied to a real image

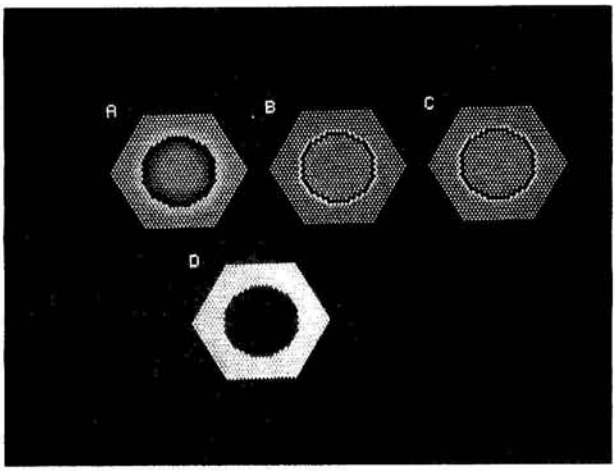

FIGURE 10: The method applied to Craik's figure.

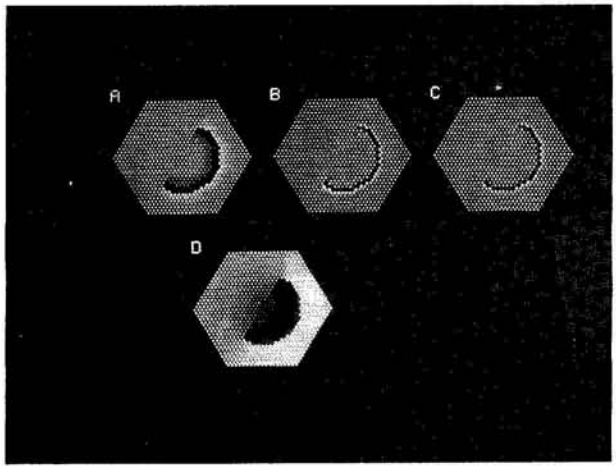

FIGURE 11: Apparent lightness predicted for incomplete figure.

The subfigures in the above have the following interpretation:

A	Input - logarithm of image intensity	p_{ij}
B	Differenced image	d_{ij}
C	Thresholded difference	t_{ij}
D	Output - computed lightness	l_{ij}
E	Illumination distribution $(p_{ij}-l_{ij})$	s_{ij}

3. Implications and Conclusions.

3.1 Parallel Image Processing Hardware.

The methods described here for forward transforming, thresholding and inverse transforming immediately tempt one to think in terms of electronic components arranged in parallel layers. Enough has been said about different models to make it clear how one might connect such components. Large scale integrated circuit technology may be useful, provided the signals are either converted from analog to digital form or better still, good linear circuits are available in this form.

Construction of such devices would be premature until further experimentation is performed to decide on optimal tesselations, optimal negative surrounds, thresholding operations and normalization schemes. These decisions are best guided by computer simulation.

3.2 Cognitive Psychology.

One of the artificial scenes was created to illustrate Craik's illusion (Brindley 1960, Cornsweet 1970). Here a sharp edge is bordered by second-order gradients. As one might expect the smooth gradients are lost in the thresholding and reconstruction produces two regions each of uniform brightness. The difference in brightness between the regions is equal to the original intensity step at the edge.

The fact that the process presented here falls prey to this

illusion is of course no proof that humans use the same mechanism. It is interesting that this technique allows one to predict for example the appearance of pictures containing incompletely closed curves with second-order gradients on either side.

3.3 Neuro-physiology.

The method described here for obtaining lightness from image intensity suggests functions for a number of structures in the primate retina. The horizontal cells appear to be involved in the forward transformation, while some of the amacrines may be involved in the inverse transformation. For details see the paper by David Marr (Marr 1974), in which he uses this hypothesis to explain an astonishing number of facts about the retina.

3.4 Conclusion.

A simple layered, parallel technique for computing lightness from image intensity has been presented. The method does not involve an ability to describe or understand the scene, relying instead on the spatial differences in the distribution of reflectance and illumination. The forward step involves accentuating the edges between regions. The output of this step is then thresholded to remove illumination gradients and noise. The inverse step merely undoes the accentuation of the edges.

Physical models have been given which can perform this computation efficiently in parallel layers of simple networks. The method has been simulated and applied to a number of images. The method grew out of an attempt to extend Land's method to two dimensions and fills the need for a lightness judging process in his retinex theory of color perception.

The possibility of processing an image in such a parallel, simple fashion without higher-level understanding of the scene reinforces my belief that such low-level processing is of importance in dealing with a number of features of images. Amongst these are shading, stereo disparity, focus, edge detection, scene segmentation and motion parallax. Some of this kind of processing may actually happen in the primate retina and visual cortex. The implications for image analysis are that it may well be that a number of such pre-processing operations should be performed automatically for the whole image to accentuate or extract certain attributes before one brings to bear the more powerful, but tedious and slow sequential goal-directed methods.

Bibliography.

Brindley, G.S. (1960) "Physiology of the Retina and Visual Pathway." (Monograph No.6 of the Physiological Society), London: Edward Arnold Ltd.

Cornsweet, T. (1970) "Visual Perception." New York: Academic Press.

Garabedian, D.R. (1964) "Partial Differential Equations." New York: John Wiley.

Goethe, J.W. von. (1810) "Zur Farbenlehre." Tuebingen.

Hardy, A.C. (Ed.) (1936) "The Handbook of Colorimetry." Cambridge, Mass: M.I.T. Press.

Helmholtz, H.L.F. (1867) "Handbuch der Physiologischen Optik." Leipzig: Voss. Also translated: Southall, J.P.C.S. "Handbook of Physiological Optics." New York: Dover Publications.

Helson, H. (1938) "Fundamental Problems in Color Vision I." Journal of Experimental Psychology, 23.

Helson, H. (1940) "Fundamental Problems in Color Vision II." Journal of Experimental Psychology, 26.

Hering, E. (1875) "Zur Lehre vom Lichtsinne - Grunzuege einer Theorie des Farbsinnes." SBK. Akad. Wiss. Wien Math. Naturwiss. K.70.

Horn, B.K.P. (1968) "The Application of Fourier Transform Methods to Image Processing." S.M. Thesis, E.E. Department, M.I.T. pp. 81-87 & pp. 93-94.

Judd, D.B. (1940) "Hue, Saturation and Lightness of Surface Colors with Chromatic Illumination ." Journal of the Optical Society of America, 30.

Judd, D.B. (1952) "Color in Business, Science and Industry." New York: John Wiley.

Land, E.H. (1959) "Experiments in Color Vision." Scientific American.

Land, E.H. (1964) "The Retinex." American Scientist, 52.

Land, E.H. & McCann, J.J. (1971) "Lightness Theory." Journal of the Optical Society, 61.

Lettvin, J.Y. (1967) "The Color of Colored Things." Quarterly Progress Report, 87, Research Laboratory for Electronics, M.I.T.

Marr, D. (1974) "An Analysis of the Primate Retina." A.I. Memo. 296, Artificial Intelligence Laboratory, M.I.T.

Maxwell, J.C. (1856) "On the Unequal Sensitivity of the Foramen Centrale to Light of Different Colours." Rep. Brit. Assoc.

Newton, Sir Isaac. (1704) "Opticks." London: Samuel Smith & Benjamin Walford. Also New York: Dover Publications.

Richards, W. (1971) "One-stage Model for Color Conversion." Journal of the Optical Society, 62.

Richtmeyer, R.D. & Morton, K.W. (1957) "Difference Methods for Initial Value Problems." New York: John Wiley.

Young, T. (1820) "On the Theory of Light and Colour." Philosophical Transactions. Also in: Teevan, R.C. & Birney, R.C. (Ed.) (1961) "Color Vision." Van Nostrand.