MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY
and
CENTER FOR BIOLOGICAL INFORMATION PROCESSING
WHITAKER COLLEGE

A.I. Memo No. 909 April 1987
C.B.I.P. Memo No. 25

LEARNING A COLOR ALGORITHM FROM EXAMPLES

A. Hurlbert and T. Poggio[t]

ABSTRACT:

We show that a color algorithm capable of separating illumination from reflectance in a Mondrian world can be learned from a set of examples. The learned algorithm is equivalent to filtering the image data – in which reflectance and illumination are intermixed – through a center-surround receptive field in individual chromatic channels. The operation resembles the "retinex" algorithm recently proposed by Edwin Land. This result is a specific instance of our earlier result that a standard regularization algorithm can be learned from examples. It illustrates that the natural constraints needed to solve a problem in inverse optics can be extracted directly from a sufficient set of input data and the corresponding solutions. The learning procedure has been implemented as a parallel algorithm on the Connection Machine System.

© Massachusetts Institute of Technology 1986

This report describes research done within the Artificial Intelligence Laboratory and Thinking Machines Corporation. Support for the A.I. 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 N00014–85–K–0124. Support for this research is also provided by a grant from the Office of Naval Research, Engineering Psychology Division and by gift of the Artificial Intelligence Center of Hughes Aircraft Corporation to T. Poggio.

[t]Thinking Machines Corporation and M.I.T. A.I. Lab

Computational vision has derived effective solutions to early vision problems such as edge detection, stereopsis and structure from motion by exploiting general constraints on the imaging process and the natural world. An important question is: how does a visual system learn the algorithms it uses? Can they – and the underlying natural constraints – be learned automatically from sets of examples? We have explored these questions for the computational problem of color vision, and implemented a parallel learning procedure on the Connection Machine System.

Color constancy points to a difficult computation underlying human color vision. We do not merely discriminate between different wavelengths of light; we assign roughly constant colors to objects even though the intensity signals they send to our eyes change as the illumination varies across space and chromatic spectrum. Perfect color constancy would result from a computation that extracts the invariant spectral reflectance properties of surfaces from the image intensity signal, in which reflectance and illumination are mixed. The fact that the colors we see are not exactly invariant suggests that our visual system performs a computation with a similar goal, but less exact results. The computation is typical of the difficult problems of inverse optics, in which the information supplied by a two-dimensional image is insufficient by itself to solve for a unique three-dimensional scene. Natural constraints must be found and applied to the problem in order to solve it.

"Retinex" lightness algorithms, pioneered by Land (Land, 1959, 1985, 1986; Land and McCann, 1971) and explored by others (Horn, 1974; Blake, 1985; Hurlbert, 1986) illustrate one successful approach to the computation. The retinex algorithms assume that the visual system performs the same computation in each of three independent chromatic channels. The algorithms further assume that in each channel, the image intensity signal, or

more precisely, the image *irradiance*, s', is proportional to the product of the illumination intensity e' and the surface spectral reflectance r' in that channel:

$$s'(x,y) = r'(x,y)e'(x,y). \tag{1}$$

This form of the intensity equation makes the implicit assumption that the irradiance s' has no specular components and that the color channels have been chosen appropriately. [1]

Retinex algorithms seek to solve Equation (1) for *lightness*, which is an approximation to $r'(x,y)$, in each channel. The resulting triplet of lightnesses labels a constant color in color space. To make a solution possible, retinex algorithms restrict their domain to a world of Mondrians, two-dimensional surfaces covered with patches of random colors (see Figure 1). The algorithms then make the explicit assumptions that (1) $r'(x,y)$ is uniform within patches but has sharp discontinuities at edges between patches and that (2) $e'(x,y)$ varies smoothly across the Mondrian.

Most retinex algorithms first take the logarithm of both sides of Equation (1), converting it to a sum:

$$s(x,y) = r(x,y) + e(x,y), \tag{2}$$

where $s = log \quad s'$, $r = log \quad r'$ and $e = log \quad e'$. The two assumptions are then exploited to break down the sum into its two components.

The most recent retinex algorithm (Land, 1986) employs an operator that divides the image irradiance of Equation (1) at each location by a weighted average of the irradiance at all locations in a large surround. The

[1] For a detailed derivation of the intensity equation, see Appendix 1.

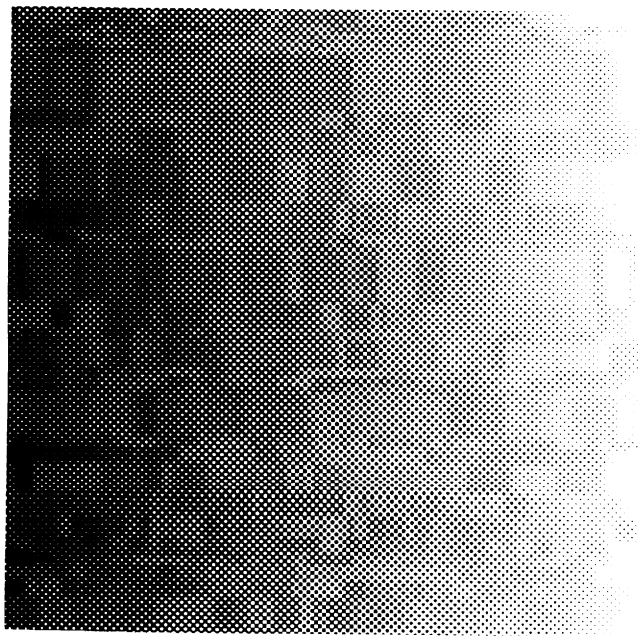

Figure 1. A Mondrian under an illumination gradient, generated by adding together two 320x320 images: one is the (log) reflectance image, an array of rectangles each with a different, uniform grey-level (see Figure 2(a)); the other is the (log) illumination image, in which the pixel values increase linearly in the same way across each row.

log of the operator's result is called lightness. The triplets calculated by the algorithm fall close to the colors humans see when viewing a Mondrian under illuminants with strong spatial gradients. The form of this operator is similar to that derived in our earlier formal analysis of the lightness problem (see Hurlbert, 1986). The main difference between the two is that the analytically derived operator takes the log of irradiance before averaging, and so is linear in the logs, whereas Land's algorithm averages before taking the log, and so is not linear in the logs. As discussed below, the numerical difference between the two results is small.

We set out to see whether a simple algorithm could learn from examples how to extract reflectance from image irradiance, and whether what it would

learn would resemble one of the above algorithms. The examples we use are pairs of images: an input image of a Mondrian under illumination that varies smoothly across space and an output array that displays the reflectance of the Mondrian separately from the illumination. We then synthesize an operator from the examples by finding the *linear* estimator that best maps the input into its two output components, using optimal linear estimation techniques.

For computational convenience we train the operator on *one-dimensional* vectors that represent horizontal scan lines across the Mondrian images (see Figure 2). We generate many different input vectors s by adding together different random r and e vectors, according to Equation (2). Each vector r represents a pattern of step changes across space, corresponding to one row of the output reflectance image (see Figure 3a). Each vector e represents a smooth gradient across space with a random offset and slope, corresponding to one row of the output illumination image. (In our implementation, we appended r to e to create an output vector twice as long as the input.) We then arrange the "training vectors" s and r as the columns of two matrices S and R, respectively. Our goal is then to compute the optimal solution L of

$$LS = R, \tag{3}$$

where L is a linear operator represented as a matrix.

It is well known that the solution of this equation that is optimal in the least squares sense is

$$L = RS^{+}, \tag{4}$$

where S^{+} is the Moore-Penrose pseudoinverse (see, for example, Albert 1972). We compute L using the technique of regularizing the pseudoinverse to

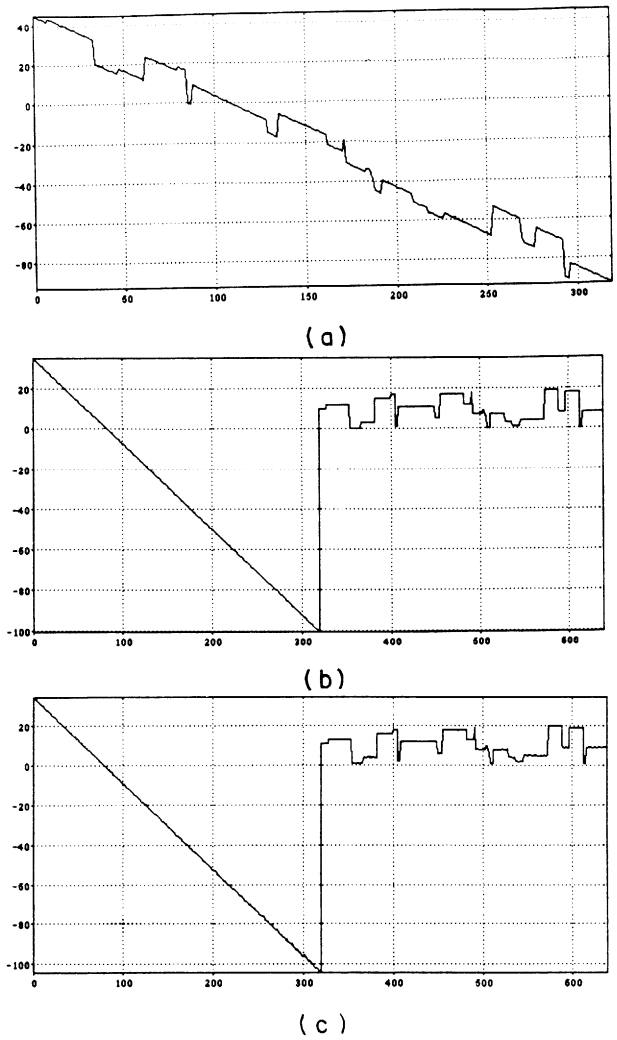

Figure 2. (a) and (b) A pair of one-dimensional examples like those used to *train* the algorithm. (a) shows the *input data*, which is a random Mondrian reflectance pattern superimposed on a linear illumination gradient with a random slope and offset. Each example can be thought of as a horizontal scan line across a Mondrian such as the one in Figure 1 (which was generated by stacking similar one-dimensional examples). Each example, 320 pixels in length, has a different reflectance pattern and a different linear illumination gradient. (b) shows the corresponding *output solution*, in which the illumination and the reflectance have been separated and concatenated. We used 1500 such pairs of input-output examples to train the operator shown in Figure 4. (c) shows the result obtained by the trained operator when it acts on the input data (a), not part of the training set. It should be compared with (b). This result is fairly typical: in some cases the prediction is even better, in others it is worse.

obviate numerical stability problems (see Appendix 3). The number of examples we use is significantly larger than the number of elements in each vector in S, in order to overconstrain the problem. Once L is computed it is tested on new scan lines, generated in the same way. The pseudoinverse may also be computed by recursive techniques which improve its form as more data become available (see Appendix 3). The latter procedure, although equivalent to our "one-shot" computing technique, may seem intuitively more like learning.

We find that the trained operator L, given a new s as input, recovers a good approximation to the correct output vectors r and e. Operating on a two-dimensional Mondrian, generated by stacking appropriately many one-dimensional s vectors, L also yields a satisfactory approximation to the correct output image (see Figure 3b).

It seems that our scheme has successfully learned an algorithm that performs the color computation correctly in a Mondrian world. *What* algorithm has been learned? What is its relationship to the filters described above? To answer these questions we examine the structure of the matrix L. We expect that because the operator should perform the same action on each point in the image, i.e. that it should be space-invariant, the central part of L should be a *convolution* matrix, in which each row is the same as the row above but displaced by one element to the right. In the peripheral parts of the matrix, this form will be corrupted by boundary effects. Inspection of the matrix and appropriate averaging of the relevant rows (see Figure 4) confirm these expectations. Like Land's psychophysically tested filter, it has a narrow positive peak and a broad, shallow, negative surround (see Figure 4) that extends beyond the range we can observe, but not over the entire image.

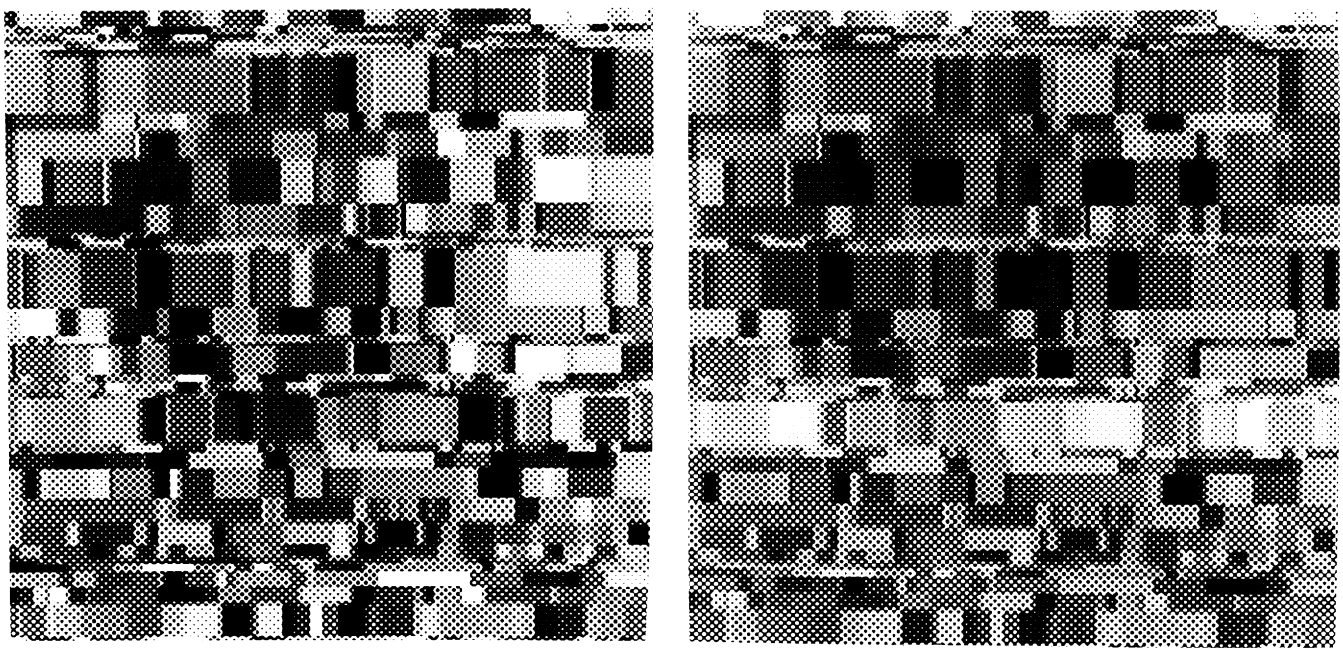

Figure 3. (a) The (log) reflectance image that is one component of the Mondrian of Figure 1. This image represents the Mondrian of Figure 1 under uniform illumination. (b) The (log) reflectance image that the trained operator produces when it acts on the Mondrian of Figure 1. The operator has been trained on a set of one-dimensional examples different from those used to generate the Mondrian of Figure 1.

Our algorithm is not exactly identical with Land's: the filter of Figure 4 subtracts from the value at each point the average value of the logs at all points in the field, rather than the log of the average values. As mentioned above, the difference between the outputs of the two filters is small in most cases, and both agree well with psychophysical results (Land, personal communication). We have explicitly compared the performance of Land's scheme on the Mondrian of Figure 1 with a scheme equivalent to our learned algorithm (filtering the log of the image through a filter like that shown in Figure 4). The resulting arrays approximate the correct outputs equally well, and are very similar to each other.

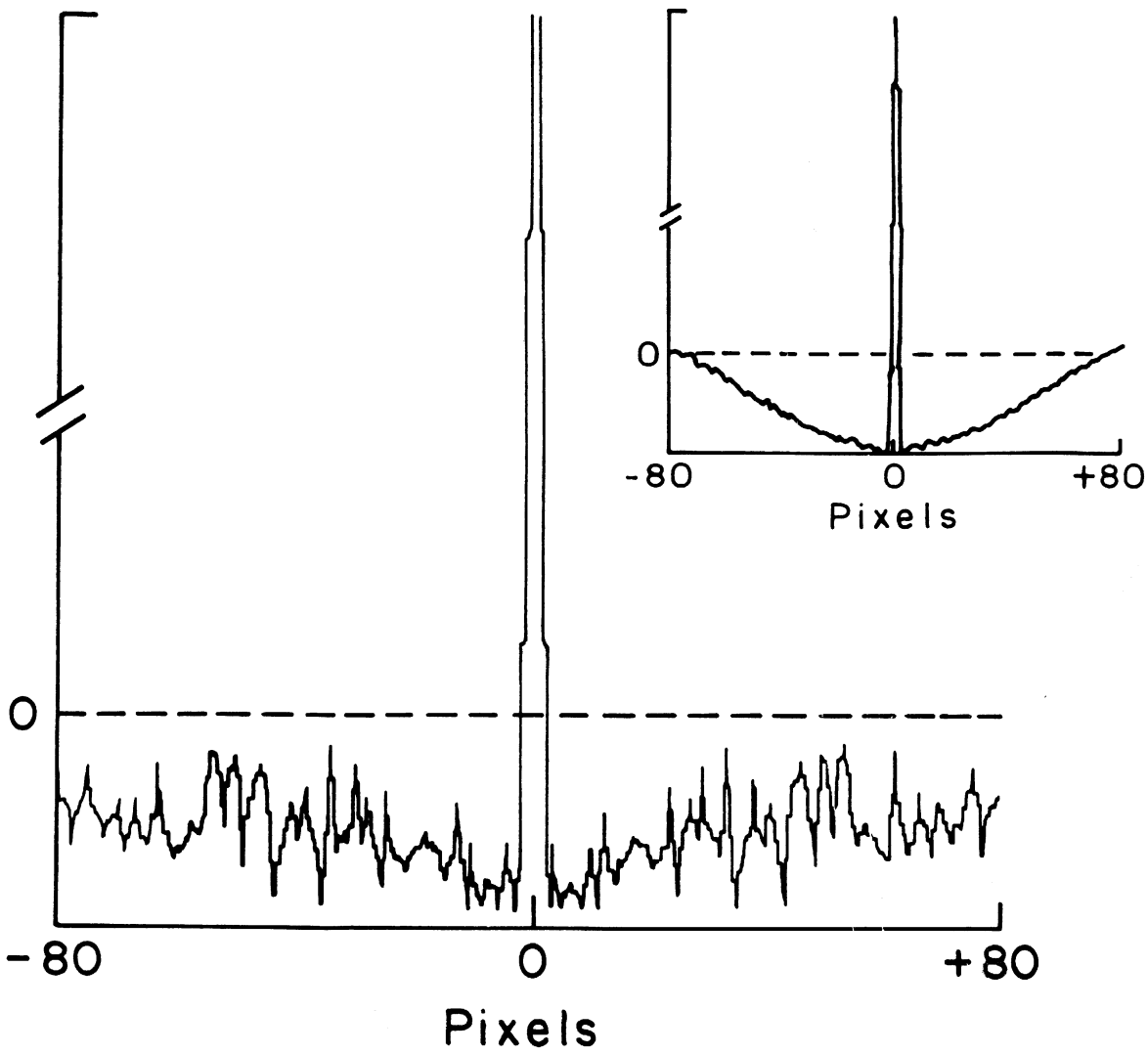

Figure 4. The learned filter, which extracts reflectance from a Mondrian image under a linear illumination gradient. The operator that is learned is a matrix that acts on one-dimensional vectors. Assuming that the operator is space invariant when boundary effects can be ignored, we estimate the shape of the corresponding filter by summating the central rows of the matrix, shifting them appropriately. The one-dimensional "receptive field" that results has a sharp positive peak and a shallow surround that extends beyond the range we can estimate reliably, which is the range we show here. The filter shown here was learned from a set of examples with linear illumination gradients (see Figure 2). When logarithmic illumination gradients are used, a qualitatively similar receptive field is obtained. In a separate experiment we used a training set of one-dimensional Mondrians with either linear illumination gradients or slowly varying sinusoidal illumination with random wavelength, phase and amplitude. The resulting filter is shown in the inset. The inhibitory surround clearly decays back to zero.

To investigate the way in which the shape of the learned operator varies with the type of illumination gradient on which it is trained, we constructed a second set of examples. In addition to vectors with random linear illumination gradients, this set contained an equal number of vectors with random sinusoidally varying illumination components. The operator trained on this mixture of illumination types differs from the operator trained on strictly linear ones in the shape of its surround. Whereas the latter has a broad negative surround that remains virtually constant throughout its extent, the new operator's surround (see Figure 4b) has a smaller extent and returns smoothly to zero from its peak negative value in its center.

The difference between the two operators illustrates an interesting feature of the learning algorithm: it adapts to its environment. The results imply that the optimal operator for images with strictly linear illumination gradients is one whose surround takes a constant average over a range smaller than the entire image. On the other hand, the surround of the optimal operator for images with smoothly varying illumination gradients is a low-pass filter that separates the illumination from the sharply-varying reflectance.

Our learning procedure is motivated by our previous observation (Poggio and Hurlbert, 1984; see also Poggio et al., 1985b) that standard regularization algorithms in early vision define linear mappings between input and output and therefore can be learned associatively under certain conditions (see Appendix 3). Our algorithm synthesizes the optimal linear operator L that maps as closely as possible, in the least squares sense, the image irradiance into its reflectance and illumination components for this class of images and illumination gradients. The technique of optimal linear estimation that it uses is closely related to optimal Bayesian estimation (see Albert, 1972 and Appendix 4).

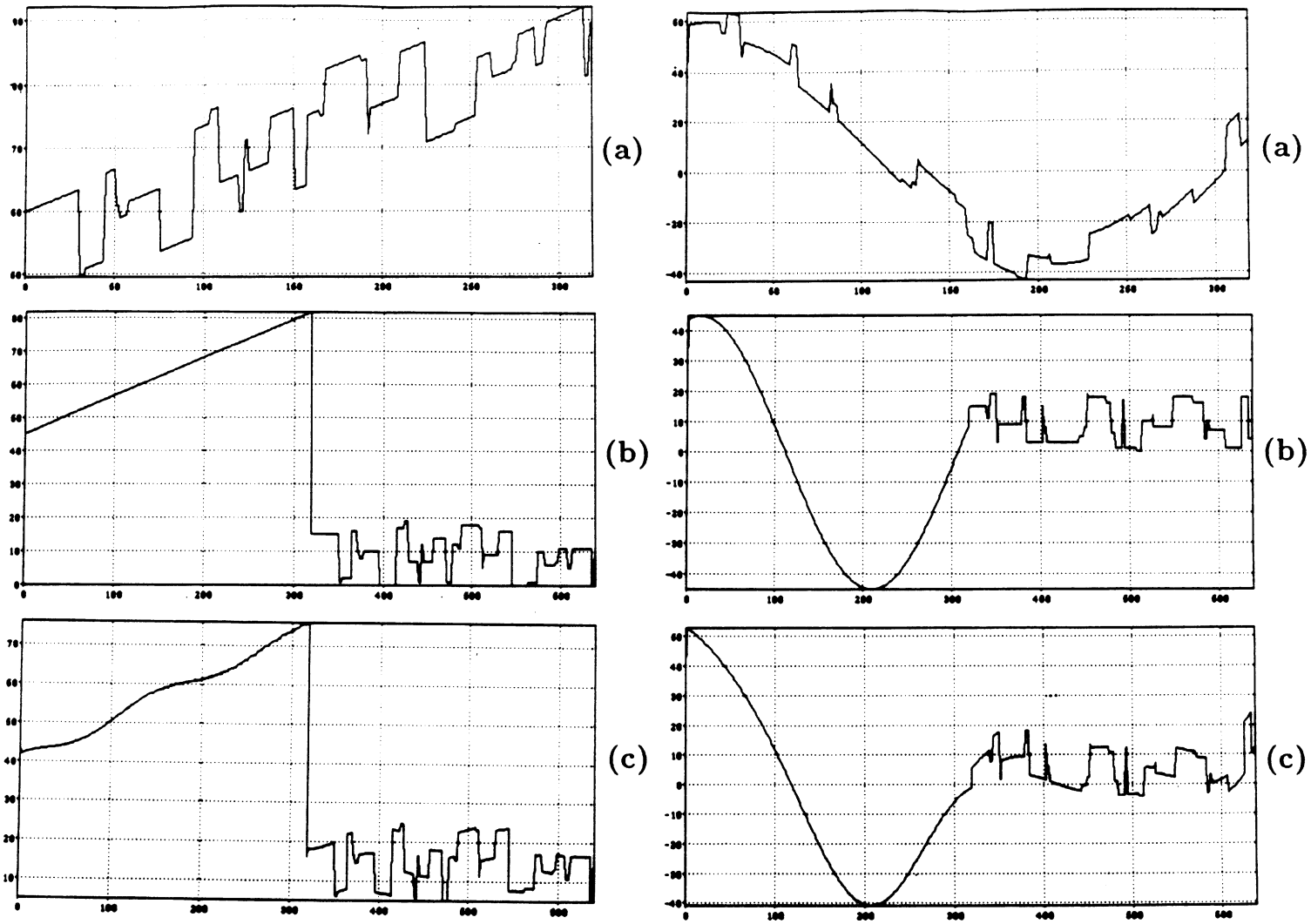

Figure 5. In a separate experiment we trained the operator on a set of 1500 one-dimensional Mondrians with either linear gradients of illumination or sinusoidally varying illumination with random wavelengths between 0.5 and 1.5 times the length of the pattern, and random phases and amplitudes. We show here how the operator performs on a new input pattern with a linear illumination gradient (left) and on a pattern with a sinusoidal illumination (right). As in Figure 2, (a) shows the *input data.* (b) shows the corresponding *output solution*, in which the illumination and the reflectance have been separated and concatenated. (c) shows the result the trained operator yields when it acts on the input data (a), not part of the training set. It should be compared with (b). The filter corresponding to this operator is shown in the inset of Figure 4.

The learned L acts on new input images to yield good approximations

to the correct output images. Note that the result of applying the matrix L to a two-dimensional Mondrian image (see Figure 3) is not as good as the one-dimensional examples suggest. The reason is that the matrix L has been learned from one-dimensional examples, and in the case of Figure 3 it has been applied to each row of the Mondrian independently. Small errors in offsets and scaling factors for each row that would only slightly affect the one-dimensional result become more obvious as they vary from row to row in the two-dimensional case.

Training of the operator on two-dimensional examples is possible, but computationally very expensive if done in the same way. The present computer simulations require several hours when run on standard serial computers (up to 20 hours on a Symbolics 3640 for 512x512 images). The two-dimensional case will need much more time. We expect that it will be feasible only on the latest version of the Connection Machine (the 65K-processor CM-2 with floating point hardware) of Thinking Machines Corporation. Our one-dimensional learning scheme runs orders of magnitude faster on a CM-1 Connection Machine System with 16K-processors. It is possible to use much more efficient methods of computing the pseudoinverse and especially approximations to it (see for example Albert, 1972 and Kohonen, 1977).

The calculation of a regularized pseudoinverse may also be implemented by parallel networks of simple processors or by analog networks that bear some resemblance to biological systems. In particular, it could be computed by so-called "neural" networks using a gradient descent method (also called back-propagation in recent papers, see Rumelhart et al., 1986) and *linear* units. Since the pseudoinverse is the best linear approximation in the L_2 norm, gradient descent minimizing the square error between the actual output and desired output of a fully connected linear network is guaranteed to

converge, albeit slowly. Thus gradient descent in weight space would give the same result we obtain, the global minimum. In terms of a network of linear units, the training scheme we have run, using one-dimensional Mondrians 512 pixels long, corresponds to learning the equivalent of up to about half a million weights.

The significance of our result lies not so much in the specific estimation technique we used, but in the form of the filter we obtain. It is qualitatively the same as that which results from the direct application of regularization methods exploiting the spatial constraints on reflectance and illumination described above (see Table 1 in Poggio et al., 1985a; Poggio and staff, 1985b; Appendix 2). The Fourier transform of the filter of Figure 4 is approximately a bandpass filter that cuts out low frequencies due to slow gradients of illumination and preserves intermediate frequencies due to step changes in reflectance. The large "inhibitory" surround also provides normalization to average grey in the field (see Hurlbert, 1986).

We do not think that our results mean that color constancy may be learned during a critical period by biological organisms. It seems more reasonable to consider them simply as a demonstration on a toy world that evolution may recover and exploit natural constraints hidden in the physics of the world. The shape of the filter in Figure 4 is suggestive of the "non-classical" receptive fields that have been found in V4, a cortical area implicated in mechanisms underlying color constancy (Desimone et. al., 1985; Wild et al., 1985; Zeki, 1983a,b).

Finally, it is important to stress that the general solution of the problem of color constancy requires much more work: real images are noisy, objects are three-dimensional, and there are shading, shadows, and specularities. We

are presently extending the simple learning techniques described here in order to deal with the full complexity of real scenes.

Appendix 1

The Intensity Equation

This appendix illustrates the transformation of the photoreceptor activity necessary in order to decompose the intensity equation into a sum of two components representing the surface reflectance and the source illumination intensity.

In most natural scenes, the light reflected from objects includes both diffuse and specular components. To simplify the intensity equation, we assume that all reflection is Lambertian, or that there are no specularities. In this case, the light reflected by a surface is solely the product of its spectral reflectance and the illumination intensity that falls on it. The amount of reflected light that reaches the eye further depends on the angles between the illumination source, the reflecting surface, and the eye, and the response of the eye to the light depends on the spectral sensitivity of its photoreceptors. We may therefore write the intensity signal registered by the eye as:

$$s'^{\nu}(x) \;=\; log \int a^{\nu}(\lambda) r'(\lambda, x) e'(\lambda, x) d\lambda, \qquad (A1.1)$$

where ν labels the spectral type of the photoreceptor ($\nu = 1, ...4$ for humans, counting 3 cone types and 1 rod type), $a^{\nu}(\lambda)$ is the spectral sensitivity of the νth-type photoreceptor and $s'^{\nu}(x)$ its activity. $r'(\lambda, x)$ is the surface spectral reflectance and $e'(\lambda, x)$ the *effective irradiance.* We group together the geometric factors influencing the intensity signal by defining the effective irradiance as the ambient illumination modified by the orientation, shape,

and location of the reflecting surface. The effective irradiance is the intensity that the surface would reflect if it were white, that is, if $r'(\lambda, x) = 1$.

Equation A1.1 is not separable into a product of reflectance and illumination components in a single color channel. To make it separable, we must make a transformation to new color channels that are combinations of the photoreceptor activities. We first choose basis functions $p^i(\lambda)$ and $q^i(\lambda)$ such that for most naturally-occurring illuminants and surface reflectances

$$\begin{aligned} e'(\lambda, x) &\approx \sum_i p^i(\lambda) e'^i(x) \\ r'(\lambda, x) &\approx \sum_i q^i(\lambda) r'^i(x) \end{aligned} \qquad (A1.2)$$

The basis transformation (for a review of the origins of this idea see Maloney, 1985; see also Buchsbaum, 1980 and Yuille, 1984) leads to the following equation

$$s'^\nu(x) = T^\nu_{i,j} e'^i(x) r'^j(x), \qquad (A1.3),$$

where the tensor T is defined as

$$T^\nu_{i,j} = \int d\lambda a^\nu(\lambda) p^i(\lambda) q^j(\lambda),$$

where $\nu = 1, ...4$ and $i, j = 1, ...N$, the p's and the q's are the basis functions for the illuminant and for the albedo, respectively, and the sum is taken over repeated indices.

To simplify further analysis, we impose the conditions that the $p^i = q^i$ and that the $p^i(\lambda)$ are orthogonal with respect to the $a^\nu(\lambda)$. This orthogonality is insured if, for example, the $p^i(\lambda)$ do not overlap with respect to λ. In the simplest case, the basis functions may be monochromatic: $p^i(\lambda) = \delta(\lambda - \lambda_i)$.

Substituting A1.2 into A1.1 then yields:

$$s'^{\nu}(x) = log \sum_i e'^{i}(x) r'^{i}(x) \int d\lambda a^{\nu}(\lambda) p^{i}(\lambda) p^{i}(\lambda) \qquad (A1.4)$$

If we define the matrix $T_i^{\nu} = \int d\lambda a^{\nu}(\lambda) p^{i}(\lambda) p^{i}(\lambda)$, where $\nu = 1, ...4$ and $i = 1, ...N$, we obtain

$$antilog \left[s'^{\nu}(x) \right] = \sum_i T_{\nu i} e'^{i}(x) r'^{i}(x) \qquad (A1.5)$$

If the $p^{i}(\lambda)$ are suitably chosen, $T_{\nu i}$ is invertible. Then the linear equations represented in A1.5 yield the following solution:

$$(T_{\nu i})^{-1} \, antilog \left[s'^{\nu}(x) \right] = e'^{i}(x) r'^{i}(x) \qquad (A1.6)$$

or,

$$s'^{i}(x) = e'^{i}(x) r'^{i}(x) \qquad (A1.7)$$

where $s'^{i}(x) = (T_{\nu i})^{-1} antilog \left[s'^{\nu}(x) \right]$. Taking logarithms of A1.7 yields

$$log \; s'^{i}(x) = log \; e'^{i}(x) \; + log \; r'^{i}(x)$$

or

$$s^{i}(x) = e^{i}(x) + r^{i}(x) \qquad i = 1, \ldots, N, \qquad (A1.8)$$

where $s^{i}(x) = log \; s'^{i}(x)$ and so on, which is the desired equation. The extension to the two-dimensional case is clear.

$s^{i}(x)$ is a linear combination of the activity of different types of photoreceptors. It is important to note that the index i labels not the color channels associated with the spectral sensitivities of the different photoreceptor types

but new channels which may be similar to the biological color-opponent channels. There is no *a priori* limit on the number of new channels formed by linear combinations, but efficiency of information transmission would require it to be close to the number of photoreceptor types.

Appendix 2
A Regularization Algorithm for Color Computation

2.1 The regularization approach

Consider the equation

$$y = Az \tag{A2.1}$$

where A is a known operator. The direct problem is to determine y given z. The inverse problem is to find z, given y – the function y becomes the "data" and z the solution. Although the direct problem is usually well-posed, the inverse problem is often *ill-posed.*

Standard regularization theories for "solving" ill-posed problems have been developed by Tikhonov (Tikhonov and Arsenin, 1977) and others. The basic idea underlying regularization techniques is to restrict the space of acceptable solutions by choosing the one solution that minimizes an appropriate functional. Among the methods that can be employed (see Poggio, et. al. 1985a), the main one is to find z that minimizes

$$\|Az - y\|^2 + \lambda\|Pz\|^2. \tag{A2.2}$$

The choice of the norm $\|\cdot\|$, usually quadratic as in Equation A2.2, and of the linear *stabilizing functional* $\|Pz\|$, is dictated by mathematical considerations, and most importantly, by a physical analysis of the generic constraints on the problem. The regularization parameter λ controls the compromise between the degree of regularization of the solution and its closeness to the data.

2.2 Regularizing the Color Computation

Equation A1.8 is impossible to solve in the absence of additional constraints: there are twice as many unknowns as equations for each x. Formally, the problem is ill-posed. Regularization techniques can be used to restrict the number of allowed solutions for $e^i(x)$ and $r^i(x)$, and thereby reduce the number of unknowns, by imposing constraints on the imaging process and the physical properties of the surfaces and the source illuminant.

One constraint that may be used is *spatial regularization* (for other constraints – *spectral regularization* and *the single source assumption* – see Hurlbert, 2001 and Poggio, 1985b). The *spatial regularization* constraint formalizes and extends the retinex assumptions that (a) $r^i(x)$ is either constant or changes sharply at boundaries between different materials, and (b) $e^i(x)$ is either constant or changes more smoothly than $r^i(x)$ across space. One retinex algorithm (Horn, 1974), for example, imposes the strong constraint (in two-dimensions) that all values of $\nabla^2 s^i(x,y)$ strictly below a fixed threshold are due to $e^i(x,y)$. The regularization assumption requires only that $e^i(x)$ vary less sharply across space than $r^i(x)$ and effectively allows the limit on the spatial variation of $e^i(x)$ to be reset for each scene.

Standard regularization techniques enforce this constraint on Equation A1.8 by requiring that its solution minimize the following variational principle:

$$\sum_i [s^i - (r^i + e^i)]^2 + \lambda[\frac{d}{dx}e^i]^2 + \beta[G * r^i + \gamma\frac{d^2}{dx^2}r^i]^2, \qquad (A2.3)$$

where G is a gaussian filter with standard deviation σ. The first term requires that the solution $(r^i + e^i)$ be close to s^i; the second term enforces the constraint that the illumination vary smoothly across space; and the third

term enforces the constraint that the reflectance vary *not* too smoothly across space.

Minimizing Equation A2.3 demonstrates that the solutions r^i and e^i may be obtained by filtering s^i through a linear filter. In the Fourier domain we derive the result:

$$\tilde{r}^i(\omega) = \frac{\lambda\omega^2}{(1+\beta e^{-\omega^2\sigma^2}+\beta\gamma\omega^2+\beta\omega^2 e^{-\omega^2\sigma^2}+\beta\gamma\omega^4)(1+\lambda\omega^2)-1}\tilde{s}^i(\omega). \tag{A2.4}$$

Note that the quadratic variational principle of standard regularization cannot enforce the spatial regularization constraint with full generality. A more general regularization scheme based on Markov random fields, which leads to standard regularization as a special case, is sketched by Poggio and staff (1985).

Appendix 3
Learning a Regularization Algorithm

3.1 Associative Learning of Standard Regularizing Operators

Minimization of the regularization principle A2.2 corresponds to determining a *regularizing* operator that acts on the input data y and produces as an output the regularized solution z. Suppose now that instead of solving for z, we are given y and its regularized solution z and we want to find the operator that effects the transformation between them. This appendix demonstrates that the regularizing operator can be synthesized by associative learning from a set of examples. The argument consists of two claims, explored in detail below. The first claim is that the regularizing operator corresponding to a quadratic variational principle is linear. The second is that any linear mapping between two vector spaces can be synthesized by an associative scheme based on the computation of the pseudoinverse of the data.

3.1.1 Linearity of the regularized solution

To show that variational principles of the form of Equation A2.2 lead to a regularized solution that is a linear transformation of the data, we start with the discrete form of Equation A2.2:

$$\|Az - y\|^2 + \lambda \|Pz\|^2, \qquad (A3.1)$$

in which z and y are vectors and A and the Tikhonov stabilizer P are matrices, A does not depend on the data, and $\|\cdot\|$ is a norm.

The minimum of this functional will occur at its unique stationary point z. To find z, we set to zero the gradient with respect to z of Equation

A3.1. The solution to the resulting Euler-Lagrange equations is the minimum vector z:

$$(A^T A + \lambda P^T P)z = A^T y. \qquad (A3.2)$$

It follows that the solution z is a linear transformation of the data y:

$$z = Ly, \qquad (A3.3)$$

where L is the linear regularizing operator. (If the problem were well-posed, the operator L would equal simply the inverse of A.) It is important to note that L may depend on the given lattice of data points.

3.1.2 Learning a linear mapping

Given that the mapping between a set of input vectors y and their regularized solutions z is linear, how do we solve for it? We start by arranging the sets of vectors in two matrices Y and Z. The problem of synthesizing the regularizing operator L is then equivalent to "solving" the following equation for L:

$$Z = LY \qquad (A3.4)$$

A general solution to this problem is given by

$$L = ZY^+, \qquad (A3.5)$$

where Y^+ is the pseudoinverse of Y. This is the solution which is most robust against errors, if Equation A3.4 admits several solutions and it is the optimal solution in the least-squares sense, if no exact solution of Equation A3.4 exists. This latter case is the one of interest to us: in order to overconstrain

the problem, and so avoid look-up table solutions, we require that the number of examples (columns of Y) be larger than the rank of the matrix L. In this case, there is no exact solution of Equation A3.4 and the matrix L is chosen instead to minimize the expression

$$M = \|LY - Z\|^2. \qquad (A3.6)$$

L may be computed directly by minimizing A3.6, which yields

$$L = ZY^T(YY^T)^{-1} \qquad (A3.7)$$

In practice, we compute L using Equation A3.7, but first regularize it by adding a stabilizing functional to obviate problems of numerical stability (Tikhonov and Arsenin, 1977).

These results show that the standard regularizing operator L (parametrized by the lattice of data points) can be synthesized without need of an explicit variational principle, if a sufficient set of correct input-output pairs is available to the system. Note that by supplying as examples the physically correct solutions z, we assume that they are identical to the regularized solutions z, and enforce both regularization and correctness on the linear operator we obtain.

3.2 Recursive estimation of L

It is of particular import for practical applications that the pseudoinverse can be computed in an adaptive way by updating it when new data become available (Albert, 1972). Consider again Equation A3.7. Assume that the matrix Y consists of $n-1$ input vectors and Z of the corresponding *correct* outputs. We rewrite Equation A3.7 as

$$L_{n-1} = Z_{n-1}Y_{n-1}^{+} \qquad (A3.8)$$

If another input-output pair y_n and z_n becomes available, we can compute L_n recursively by

$$L_n = L_{n-1} + (z_n - L_{n-1}y_n)t_n^T, \qquad (A3.9)$$

where

$$t_n^T = \frac{y_n^T(Y_{n-1}Y_{n-1}^T)^{-1}}{1 + y_n^T(Y_{n-1}Y_{n-1}^T)^{-1}y_n}, \qquad (A3.10)$$

provided that $(Y_{n-1}Y_{n-1}^T)^{-1}$ exists (i.e., that the number of columns in Y is greater than or equal to the dimension of y). The case in which $(Y_{n-1}Y_{n-1}^T)^{-1}$ does not exist is discussed together with more general results in Albert (1972). Note that $(z_n - L_{n-1}y_n)$ in the updating Equation A3.9 is the error between the desired output and the predicted one, in terms of the current L. The coefficient t_n is the weight of the correction: with the value given by Equation A3.10 the correction is optimal and cannot be improved by any iteration without new data. A different value of the coefficient is suboptimal but may be used to converge to the optimal solution by successive iterations of Equation A3.10 using the same data.

Appendix 4
Optimal Linear Estimation, Regression and Bayesian Estimation

The optimal linear estimation scheme we have described is closely related to a special case of Bayesian estimation in which the best linear unbiased estimator (BLUE) is found. Consider Equation A3.3: the problem is to construct an operator L that provides the best estimation Ly of z. We assume that the vectors y and z are sample sequences of gaussian stochastic processes with, for simplicity, zero mean. Under these conditions the processes are fully specified by their correlation functions

$$E[yy^T] = C_{yy}, \quad E[zy^T] = C_{zy} \tag{A4.1}$$

where E indicates the expected value. The BLUE of z (see Albert, 1972) is, given y,

$$z^{est} = C_{zy}C_{yy}^{-1}y, \tag{A4.2}$$

which is to be compared with the regression equation

$$Ly = ZY^T(YY^T)^{-1}y. \tag{A4.3}$$

The quantities ZY^T and YY^T are approximations to C_{zy} and C_{yy}, respectively, since the quantities are estimated over a finite number of observations (the training examples). Thus there is a direct relation between BLUEs and optimal linear estimation. The learned operator captures the stochastic regularities of the input and output signals. Note that if the input vectors y are orthonormal, then $L = ZY^T$ and the problem reduces to constructing

a simple correlation memory of the holographic type (see Poggio, 1975a,b). Under no restrictions on the vectors y, the correlation matrix ZY^T may still be considered as a low-order approximation to the optimal operator (see Kohonen, 1977).

Acknowledgements

We are grateful to Edwin Land, Michael Burns, Robert Savoy, Ellen Hildreth, Jim Little, Frank Wilczek and Danny Hillis for reading the draft and for useful discussions. Alan Rottenberg developed the routines for matrix operations that we used on the Connection Machine.

Reading List

Albert, Arthur (1972). **Regression and the Moore-Penrose Pseudoinverse**, Academic Press, New York.

Blake, Andrew (1985). "On lightness computation in Mondrian world," **Central and Peripheral Mechanisms of Colour Vision**, T. Ottoson and S. Zeki, eds., Macmillan, New York, 45-59.

Buchsbaum, G. (1980) "A spatial processor model for object color perception," *J. Franklin Inst.*, 310.

Desimone, R., Schein, S.J., Moran, J. and Ungerleider, L.G. (1985). "Contour, color and shape analysis beyond the striate cortex," *Vision Res.*, **25**, 441-452.

von Helmholtz, H.L.F. (1909/1962). **Handbuch der Physiologischen Optik**, Vol. 2, J.P. Southall, ed., Dover Publications.

Horn, Berthold K.P. (1974)."Determining lightness from an image," *Computer Graphics and Image Processing*, **3**, 277-299.

Hurlbert, A. (2001). "Color computation in the visual system," *Massachusetts Institute of Technology Artificial Intelligence Laboratory Memo 814*, Cambridge, MA, in press.

Hurlbert, A. (1986). "Formal connections between lightness algorithms," *J. Opt. Soc. Am. A*, **3**, October 1986, 1684-1693.

Kohonen, T. (1977). **Associative Memory**, Springer Verlag, Heidelberg.

Land, Edwin H. (1959). "Color vision and the natural image", *Proc. Natl. Acad. Sci.*, **45**, 115-129.

Land, Edwin H. and McCann, J.J. (1971). "Lightness and retinex theory", *J. Opt. Soc. Am.*, **61**, 1-11.

Land, Edwin H. (1985). "Recent advances in retinex theory," *Central and Peripheral Mechanisms of Colour Vision*, T. Ottoson and S. Zeki, eds., Macmillan, New York, 5-17.

Land, Edwin H. (1986). "An alternative technique for the computation of the designator in the retinex theory of color vision," *Proc. Nat. Acad. Sci.*, **83**, 3078-3080.

Maloney, Laurence T. (1985) "Computational approaches to color constancy" Stanford University Tech. Report 1985-01.

Poggio,T. (1975a) 'On optimal nonlinear associative recall," *Biol. Cyber.*, **19**, 201-209.

Poggio,T. (1975b) "On optimal discrete estimation," In:*Proceedings of the First Symposium on Testing and Identification of Nonlinear Systems*, G.D. McCann and P.Z. Marmarelis (eds.), 30-37.

Poggio, T. and Hurlbert, A. (1984). "Associative learning of standard regularizing operators in early vision," *Massachusetts Institute of Technology Artificial Intelligence Laboratory Working Paper 264*, Cambridge, MA.

Poggio, T., Torre, V. and Koch, C. (1985a). "Computational vision and regularization theory," *Nature*, **317**, 314-319.

Poggio, T. and staff (1985b). "MIT Progress in Understanding Images", *Proceedings Image Understanding Workshop*, L. Baumann, ed., SAIC, McLean, VA, 25-39.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J., (1986). "Learning representations by back-propagating errors," *Nature*, **323**, 533-536.

Tikhonov, A.N. and Arsenin, V.Y. (1977). **Solutions of Ill-Posed Problems**, Winston & Sons, Washington, DC.

Wild, H.M., Butler, S.R., Carden, D., and Kulikowski, J.J. (1985). "Primate cortical area V4 important for colour constancy but not wavelength discrimination," *Nature*, **313**, 133–135.

Yuille, Alan L. (1984) "A method for computing spectral reflectance," Massachusetts Institute of Technology Artificial Intelligence Laboratory Memo 752.

Zeki, S. M. (1983a). "Colour coding in the cerebral cortex: The reaction of cells in monkey visual cortex to wavelengths and colours," *Neuroscience*, **9**, 741–765.

Zeki, S. M. (1983b). "Colour coding in the cerebral cortex: The responses of wavelength-selective and colour-coded cells in monkey visual cortex to changes in wavelength composition," *Neuroscience*, **9**, 767-781.

CS-TR Scanning Project
Document Control Form

Date : 10/18/95

Report # AIM-909

Each of the following should be identified by a checkmark:
Originating Department:

☒ Artificial Intellegence Laboratory (AI)
☐ Laboratory for Computer Science (LCS)

Document Type:

☐ Technical Report (TR) ☒ Technical Memo (TM)
☐ Other:_____

Document Information

Number of pages: 30 (36-IMAGES)
Not to include DOD forms, printer intstructions, etc... original pages only.

Originals are:
☒ Single-sided or
☐ Double-sided

Intended to be printed as :
☐ Single-sided or
☒ Double-sided

Print type:
☐ Typewriter ☐ Offset Press ☒ Laser Print
☐ InkJet Printer ☐ Unknown ☐ Other:_____

Check each if included with document:

☒ DOD Form (2) ☐ Funding Agent Form ☐ Cover Page
☐ Spine ☐ Printers Notes ☐ Photo negatives
☐ Other: _____

Page Data:

Blank Pages (by page number): _____

Photographs/Tonal Material (by page number): 4,8

Other (note description/page number):

Description :	Page Number:
(A) IMAGE MAPS (1-30) PAGES #'ED 1-30 (31-36) SCANCONTROL, DOD(2), TRGT'S (3)	
(B) CUT & PASTE FIG'S ON PAGES	4, 6, 8-9, 11

Scanning Agent Signoff:

Date Received: 10/18/95 Date Scanned: 10/24/95 Date Returned: 10/26/95

Scanning Agent Signature: Michael W. Cook

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 909	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER AD-A184385
4. TITLE (and Subtitle) Learning a Color Algorithm from Examples		5. TYPE OF REPORT & PERIOD COVERED AI Memo 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Anya Hurlbert and Tomaso Poggio		8. CONTRACT OR GRANT NUMBER(s) N00014-85-K-0124
9. PERFORMING ORGANIZATION NAME AND ADDRESS Artificial Intelligence Laboratory 545 Technology Square Cambridge, MA 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, VA 22209		12. REPORT DATE June 1987 13. NUMBER OF PAGES 30
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Office of Naval Research Information Systems Arlington, VA 22217		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Computer vision	Regularization
Color constancy	Optimal estimation
Learning	Pseudoinverse

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We show that a color algorithm capable of separating illumination from reflectance in a Mondrian world can be learned from a set of examples. The learned algorithm is equivalent to filtering the image data -- in which reflectance and illumination are mixed -- through a center-surround receptive field in individual chromatic channels. The operation resembles the "retinex" algorithm recently proposed by Edwin Land. This result is a specific instance of our earlier results that a standard regularization algorithm can be learned from

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: 909	2. GOVT ACCESSION NO.
	3. RECIPIENT'S CATALOG NUMBER: AD-A184385
4. TITLE (and Subtitle): Learning a Color Algorithm from Examples	5. TYPE OF REPORT & PERIOD COVERED: AI Memo
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Anya Hurlbert and Tomaso Poggio	8. CONTRACT OR GRANT NUMBER(s): N00014-85-K-0214
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Artificial Intelligence Laboratory, 545 Technology Square, Cambridge, MA 02139	10. PROGRAM ELEMENT PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Advanced Research Projects Agency, 1400 Wilson Blvd., Arlington, VA 22209	12. REPORT DATE: June 1987
	13. NUMBER OF PAGES: 30
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Office of Naval Research, Information Systems, Arlington, VA 22217	15. SECURITY CLASS (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Computer vision	Regularization
Color constancy	Optimal estimation
Learning	Pseudoinverse

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We show that a color algorithm capable of separating illumination from reference in a Mondrian world can be learned from a set of examples. The learned algorithm is equivalent to filtering the image data in the [illegible] and illumination are mixed [illegible] divident chromatic channels. The operation resembles [illegible] recently proposed by Edwin Land. This result is a specific [illegible] earlier results that a standard regularization algorithm can be learned from

Abstract (cont'd)

examples

It illustrates that the natural constraints needed to solve a problem in inverse optics can be extracted directly from a sufficient set of input data and the corresponding solutions. The learning procedure has been implemented as a parallel algorithm on the Connection Machine System.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Scanning Agent Identification Target

Scanning of this document was supported in part by the **Corporation for National Research Initiatives,** using funds from the **Advanced Research Projects Agency** of the **United states Government** under Grant: **MDA972-92-J1029.**

The scanning agent for this project was the **Document Services** department of the **M.I.T Libraries.** Technical support for this project was also provided by the **M.I.T. Laboratory for Computer Sciences.**

Scanned

Date: 10/24/1995

M.I.T. Libraries
Document Services