International Journal of Computer Vision, 5:1, 5-36 (1990)


# A Novel Algorithm for Color Constancy

D.A. FORSYTH*
*Robotics Research Group, Department of Engineering Science, Oxford University, England OXI 3PJ*

**Abstract**
Color constancy is the skill by which it is possible to tell the color of an object even under a colored light. I interpret the color of an object as its color under a fixed canonical light, rather than as a surface reflectance function. This leads to an analysis that shows two distinct sets of circumstances under which color constancy is possible. In this framework, color constancy requires estimating the illuminant under which the image was taken. The estimate is then used to choose one of a set of linear maps, which is applied to the image to yield a color descriptor at each point. This set of maps is computed in advance.

The illuminant can be estimated using image measurements alone, because, given a number of weak assumptions detailed in the text, the color of the illuminant is constrained by the colors observed in the image. This constraint arises from the fact that surfaces can reflect no more light than is cast on them. For example, if one observes a patch that excites the red receptor strongly, the illuminant cannot have been deep blue.

Two algorithms are possible using this constraint, corresponding to different assumptions about the world. The first algorithm, **Crule** will work for any surface reflectance. **Crule** corresponds to a form of coefficient rule, but obtains the coefficients by using constraints on illuminant color. The set of illuminants for which **Crule** will be successful depends strongly on the choice of photoreceptors: for narrowband photoreceptors, **Crule** will work in an unrestricted world. The second algorithm, **Mwext**, requires that both surface reflectances and illuminants be chosen from finite dimensional spaces; but under these restrictive conditions it can recover a large number of parameters in the illuminant, and is not an attractive model of human color constancy.

**Crule** has been tested on real images of Mondriaans, and works well. I show results for **Crule** and for the Retinex algorithm of Land (Land 1971; Land 1983; Land 1985) operating on a number of real images. The experimental work shows that for good constancy, a color constancy system will need to adjust the gain of the receptors it employs in a fashion analagous to adaptation in humans.

## 1 Introduction

People experience color as a surface property that is largely unaffected by the color of the illuminating light. This phenomenon is known as color constancy. The mechanisms of human color constancy and the circumstances under which these mechanisms work are the subjects of extensive debate. Color constancy is a valuable skill for machine vision, because it allows statements about surface properties that are invariant to viewing conditions. There is a wide and active literature on color constancy in humans and in machines, which is extensively reviewed in (Forsyth 1989b).

*The author acknowledges the support of the Rhodes Trust and of Magdalen College, Oxford.

Apparently the oldest algorithm for color constancy is the coefficient rule, which is normally attributed to von Kries (von Kries 1878; see also West and Brill 1982; Worthey 1985; Worthey and Brill 1986). This algorithm adjusts the gain of each class of photoreceptor (for example, the red, green, and blue channels in a color camera) independently to obtain surface color descriptors. The factors by which the gains are adjusted are called the coefficients. Different authors use different factors: for example, Brill and West (Brill and West 1981) divide the output of each photoreceptor class by its output for a surface patch known to be white, and Land (Land and McCann 1971) chooses coefficients such that the geometric average of photoreceptor responses is constant for each class (Brainard and Wandell 1986).

Adjusting gains is equivalent to regarding the camera output at each point in the image as a vector with one entry for each photoreceptor class, and multiplying this vector by a diagonal matrix, where the diagonal elements are the coefficients, which do not change from point to point.

The coefficient rule can fail when the spectral sensitivities of the photoreceptor classes are not disjoint because in this case the photoreceptor measurements are not independent. Land (Land 1983) has observed that if some linear combination of photoreceptor spectral sensitivities is independent, it is possible to adjust this linear combination instead. To achieve this adjustment, one multiplies the camera output vector by a constant matrix (say $\mathbf{A}$) which changes the basis to that in which the outputs are independent, by a diagonal matrix (say $\Lambda$), and then by $\mathbf{A}^{-1}$. The diagonal terms in $\Lambda$ can again be chosen in a number of different ways.

Recent research, for example (Sallstrom 1973; Buchsbaum 1980; Brill and West 1981; Brainard and Wandell 1986; D'Zmura and Lennie 1986; Maloney 1986; Maloney and Wandell 1986; Wandell 1987; Gershon 1988; and Ho et al. 1988), has modeled surface reflectances by a finite dimensional space of functions and illuminants by another finite dimensional space of functions. Surface reflectances and illuminants can then each be represented by a finite coefficient vector, and the imaging process is an algebraic interaction between the surface reflectance coefficients and the illuminant coefficients. The richness and complexity of these interactions means that the coefficient rule is insufficient to model them. This has led to a belief that the coefficient rule is an inappropriate technique for color constancy (Brainard and Wandell 1986). Despite this belief, this paper demonstrates that the coefficient rule has much to recommend it.

Given a surface with a known illuminant, it is possible to measure only a finite number of properties of its surface reflectance—at most, one property for each class of photoreceptor. As a result, it is neither correct nor helpful to see color constancy as a problem of measuring surface reflectance. Rather, color constancy programs attempt to provide surface color descriptors that are unaffected by changes in the illuminant, and are not trivial (e.g., are always fixed at some value for every surface under every light). In this article, I use the appearance of a patch under a canonical illuminant as a descriptor. According to this approach, color constancy involves predicting what an image would have looked like, had it been taken under the canonical illuminant. This requires determining what illuminant was in fact used. The illuminant information in turn specifies a map, taken from a set prepared in advance, which is applied to the image. The output of this map is the predicted appearance of the image under the canonical illuminant—that is, the color descriptors.

First, I use this approach to determine under what circumstances color constancy is possible. Then, to determine the illuminant, I use the observation that only a limited set of photoreceptor responses is possible under any given illuminant, because surfaces can reflect no more light than is cast on them. This leads to a constraint on the illuminant, because observing a photoreceptor response is prima facie evidence that certain illuminants were not employed. For example, one will not observe a strong response from a red photoreceptor under a blue light, because to do so would require that some surface reflect more red light than is actually falling on it. In fact, if one observes a strong response from a red photoreceptor, then the illuminant cannot be blue.

Two algorithms result from this analysis, distinguished by the assumptions necessary for them to work. **Crule** is a simple algorithm that works for an infinite dimensional set of surface reflectances, given sufficiently strongly controlled illuminants. **Crule** forms descriptors using the coefficient rule, with a novel technique for choosing coefficients. A more complex algorithm, **Mwext**, arises in the case of finite dimensional sets of surface reflectances. Restricting surface reflectances to a finite dimensional set means that the world is strongly constrained, so that this algorithm is able to recover a large number of parameters in the illuminant. Although **Crule** requires either a strong constraint on the illuminant or particular photoreceptor properties to succeed, **Mwext** requires unusual properties of both the illuminant and of surface reflectances.

## 2 Preliminary Assumptions

Many effects can confound color constancy in the real world. Surfaces receive light from other surfaces as well as from the primary source, so that the effective illuminant seen by one surface may differ considerably from that seen by another. Shadows create an ambiguity between one patch, half of which is in shadow, and two

patches, one of which is darker than the other. If surfaces are not at a fixed orientation to the light source, it is not always possible to determine whether a patch is dark, or is oriented so that it receives little light.

No present programs can resolve these ambiguities. Very little is known about how they could be resolved. Research on color constancy is commonly confined to considering a Mondriaan world, a world of flat, frontally presented collages of colored paper, where these ambiguities do not occur. In this article, I consider only the Mondriaan world, because, although the problem of color constancy in this world is only marginally relevant to building machines that can see, solving this problem is necessary before it will be possible to build sophisticated color constancy programs.

A number of preliminary assumptions are necessary. These are detailed here.

1. All surfaces are flat, and frontally presented and there are no significant mutual illumination effects. No shadowing occurs.
2. There is a single illuminant, which does not vary over space.
3. All surfaces are Lambertian, and all reflection is diffuse. A surface's albedo varies with wavelength only, and this variation is described by a surface reflectance function. Surfaces do not fluoresce.
4. Color constancy consists of two distinct problems. Firstly, the illuminant must be estimated in some way. Secondly, some statement about the properties of imaged surfaces must be obtained from that estimate and from the responses of the receptors.
5. The product of any surface reflectance function, any illuminant function, and any photoreceptor sensitivity can be integrated with respect to wavelength. Surface reflectance functions are never greater than one, nor less than zero. These are very weak assumptions.

Many of these assumptions are violated in practice. Assumption 1 clearly does not apply to the real world. A body of work exists that attempts to relax assumption 2. This is discussed briefly in section 7. The wings of most butterflies and many birds, for example, violate assumption 3, because their color comes from interference effects. Although surfaces are often assumed Lambertian in computer vision, there is very little information about the validity of this assumption in general—for a clear presentation of the available evidence, see (Brelstaff 1988). Some recent work has used specular components of reflected light to measure the illuminant color (Lee 1986; Klinker et al. 1987); the idea is an old one, and earlier work is described in (Judd 1960). This work encounters difficulties in the dynamic range of most cameras (when the aperature is sufficiently narrow that a specularity does not saturate the camera, it is often difficult to determine the color of objects to more than two or three bits), and with metals. Furthermore, it is not always obvious how surface color can be recovered when one knows the illuminant color.

Further assumptions that constrain properties of the illuminant will become necessary as the analysis proceeds. Although their precise statement requires the analysis of the following section, and is deferred until that section, the meaning of the assumptions is intuitively reasonable and will be stated here.

- Illuminants are "reasonable." For example, it is unreasonable to expect that one could predict the appearance under white light of a surface that one has seen under monochromatic light. Furthermore, it must be possible to describe each member of the set of illuminants that one observes (e.g., with a parametrization).
- The photoreceptors are "reasonable," given the surfaces and the illuminants. For example, it is unreasonable to expect nontrivial color constancy when the illuminants do not excite the photoreceptors.
- Given an illuminant for which two patches are metamers (that is, evoke the same receptor responses), a color constancy algorithm need not predict that they will look different under a different light. Metamers are common in theory, because there are many degrees of freedom in illuminant and surface reflectance. In practice, physical constraints on surface reflectances suggest that metamerism is uncommon (Stiles et al. 1977; Maloney 1986).
- There is a variety of different colored objects in the image. The colors of these objects are not "unreasonably" distributed (that is, the scene does not consist of only shades of some color, for example), and there are "sufficient" different colors. Neither term can be made precise at this stage. This assumption is necessary because the constraints on the illuminant become very weak when there are few colors, and the process of estimating the illuminant is then prone to error.
- For any illuminant, a reasonable measurement of the photoreceptor outputs is possible. This is a property

of the sensor quantization system. Without this assumption, an illuminant could be chosen that is so deeply colored that some photoreceptor classes have poor effective resolution, because they are excited weakly or not at all by the illuminant.

These assumptions are in practice strong. It is not always possible to ensure that images contain a wide variety of surface colors. Metamerism does occur. It is not always possible to control the surface reflectances and illuminants such that color constancy is possible. Further work is necessary to relax all these assumptions.

## 3 The Color Constancy Equation

In general in what follows, vector quantities are denoted by underlining, and vector components by subscripting.

| | |
|---|---|
| $E$ | The set of observably distinct illuminant functions, parametrized by some parameter vector $\underline{t}$. |
| $\lambda$ | The wavelength parameter. |
| $e(\lambda; \underline{t})$ | The illuminant function corresponding to the parameter $\underline{t}$. |
| $e(\lambda; \underline{t}^c)$ | The canonical illuminant. |
| $S$ | A set of surface reflectance functions. |
| $s(\lambda)$ | A surface reflectance. |
| $L$ | The number of receptors. |
| $\rho_k(\lambda)$ | The $k$th receptor sensitivity, where $k = 0, \ldots, L-1$. |
| $\Phi_l(\lambda; \underline{t})$ | $\Phi_l(\lambda; \underline{t}) = \rho_l(\lambda)e(\lambda; \underline{t})$, where $l = 0, \ldots, L-1$. This apparently arbitrary term clarifies the development of the theory. |
| $\phi_m(\lambda)$ | A member of an orthogonal basis for the space spanned by all the $\Phi_l(\lambda; \underline{t})$, where $m = 0, \ldots, L-1$. |
| $\underline{\Phi}(\lambda; \underline{t})$ | The vector whose components are $\Phi_l(\lambda; \underline{t})$, $l = 0, \ldots, L-1$. |
| $\underline{\Psi}(\cdot; \underline{t})$ | A function chosen from a set of functions, parametrized by $\underline{t}$ which takes the appearance of a surface under the canonical illuminant to its appearance under the light represented by $\underline{t}$. |

Two illuminants are "observably distinct" when, if one images all possible surface reflectances under these illuminants, one obtains two gamuts where are not exactly the same set. The response of the $k$th photoreceptor, viewing a surface reflectance $s(\lambda)$ with an illuminant $e(\lambda; t)$ is

$$\int \rho_k(\lambda)e(\lambda; \underline{t})s(\lambda)\, d\lambda$$

where the integral is over some appropriate range of wavelengths. For linear receptors such as CCD cameras, observing a very broad class of surfaces without specular reflection, this equation will correctly represent the output of the $k$th receptor. The position is more complex with human receptors, not the least because there is reason to believe that the sensitivity of receptors itself varies with the brightness and duration of exposure of a stimulus (see, for example, Bartleson 1977; Wyszecki and Stiles 1980), but at present I consider a situation where receptor sensitivity is a function of wavelength alone. This is a reasonable model for a CCD camera, and any departure from this model is likely to introduce unmanageable difficulties. I assume that the $L$ classes of photoreceptors in the camera system have different, linearly independent spectral sensitivies. If this is not the case, some of the receptors are redundant and can be ignored. It will appear from the analysis below that the choice of receptor spectral sensitivities can strongly affect color constancy. The illuminant parameter is written as a vector, although it is not obvious that it is a vector. Two cases will emerge from the analysis below. In the first case, the set of illuminants is a finite-dimensional vector space. In the second case, it is a smooth manifold.

The color of a surface will be described by its appearance under $e(\lambda; \underline{t}^c)$, the canonical illuminant. Note that many different surface reflectances cause the same receptor responses under the canonical illuminant. Thus, each descriptor refers to an equivalence class of surface reflectances. Color constancy then involves taking an image of a set of surfaces illuminated by some member of $E$, $e(\lambda; \underline{t})$, and predicting its appearance had it been illuminated by $e(\lambda; \underline{t}^c)$. It is possible to do this by constructing a function $\underline{\Psi}$ that predicts the receptor responses generated by some surface under colored lights, given the lighting parameter $\underline{t}$ and the receptor responses generated by that surface under the canonical illuminant. Assume that the illuminant is known, and is given by $\underline{t}$. We then have

$$\underline{\Psi}\left(\int \underline{\Phi}(\lambda; \underline{t}^c)s(\lambda)\, d\lambda; \underline{t}\right) = \int \underline{\Phi}(\lambda; \underline{t})s(\lambda)\, d\lambda \quad (1)$$

for all $s(\lambda)$. In this way an operation on surface color is associated with an illuminant.

Given this function, the descriptor for a patch in an image taken under a known illuminant, whose parameter

is $\underline{t}$, can be found by computing its preimage in $\underline{\Psi}(\cdot; \underline{t})$. Figure 1 shows a gamut observed under white light. If $\underline{t}$ represented green or blue light, $\underline{\Psi}(\cdot; \underline{t})$ would skew the gamut to those of figures 2 or 3, respectively. If $\underline{\Psi}(\cdot; \underline{t})$ is bijective, then there is a single possible descriptor for that patch. If not, some number of degrees of freedom in the descriptor will remain unconstrained. This case is easily detected in practice, because the gamut of receptor responses will have fewer than the maximum number of degrees of freedom.[1] This means that one or more of the photoreceptors is redundant. In this case if one ignores the redundant photoreceptors, the analysis below passes through (it will emerge that $\underline{\Psi}$ is linear, so that no intersecting singular behavior is possible when it is not bijective). Since there are many different possible conventions for managing the unmeasured degrees of freedom, and because it does not affect the analysis, I do not deal with this case further.

Now consider an orthonormal basis $\{\phi_m(\lambda), 0 \leq m \leq L - 1\}$ for the space spanned by the component functions $\Phi_l(\lambda; \underline{t}^c)$ of $\underline{\Phi}(\lambda; \underline{t}^c)$. Without loss of generality, assume that the $L$ functions $\Phi_l(\lambda; \underline{t}^c)$ are linearly independent. If they are not, either the receptor spectral

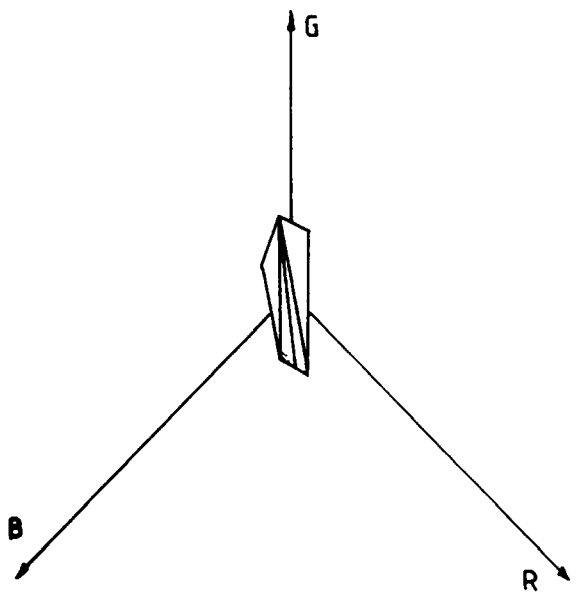



*Fig. 2.* The convex hull of the gamut observed for a Mondriaan image under green light. Notice that there are significant differences between this gamut, and that of figures 1 and 3. The feasible set consists of maps taking this gamut to a subset of the gamut of figure 1.

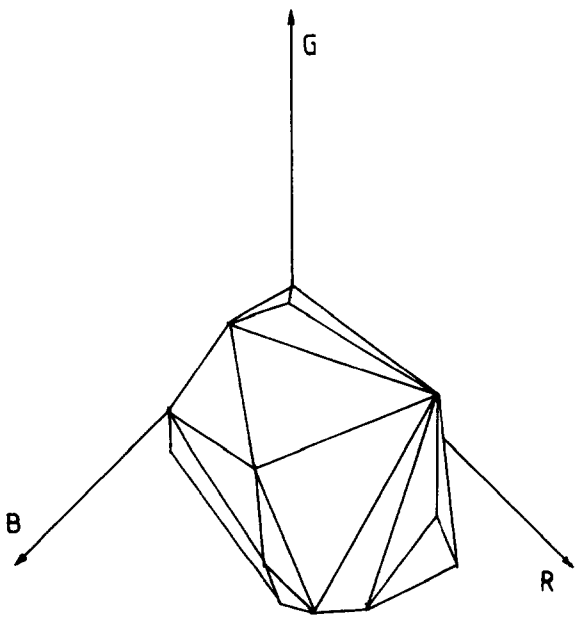



*Fig. 1.* The convex hull of the gamut obtained by observing 180 color chips under white light. This formed the observed canonical gamut for the experimental work. If $\underline{t}$ represents the green light for the gamut shown in figure 2, $\underline{\Psi}(\cdot; \underline{t})$ would take this set to a superset of that gamut.

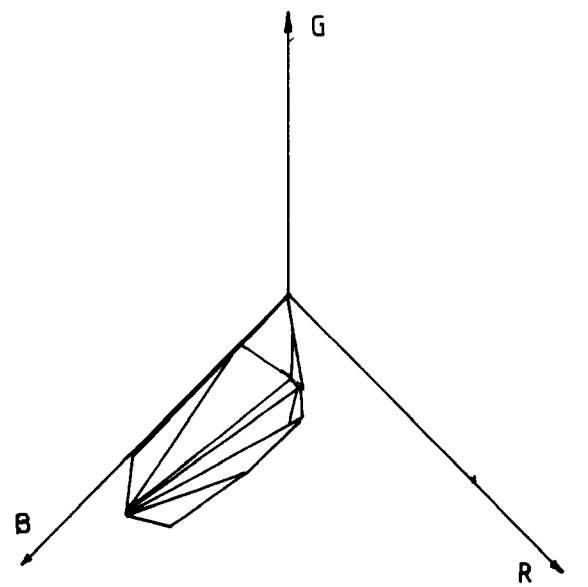



*Fig. 3.* The convex hull of the gamut observed for a Mondriaan image under blue light. Notice that there are significant differences between this gamut, and those of figures 1 and 2. **Crule** uses this skewing effect to infer possible illuminants.

sensitivities are not linearly independent or the canonical illuminant has been poorly chosen. It is clear that such an orthonormal basis exists (apply the Gramm-Schmidt algorithm; see, for example, Eisenschitz 1966). Then there is a *unique* decomposition of both $\Phi_l(\lambda; \underline{t}^c)$, and of $\Phi_l(\lambda; \underline{t})$ in terms of this orthonormal basis: say,

$$\Phi_l(\lambda; \underline{t}^c) = \sum_{j=0}^{j=L-1} a_{lj}\phi_j(\lambda)$$

Since this basis $\{\phi_m(\lambda), 0 \leq m \leq L\}$ was introduced to span the space spanned by the functions $\Phi_l(\lambda; \underline{t}^c)$, it may be the case that $\Phi_l(\lambda; \underline{t})$ does not belong to $Span\{\phi_m(\lambda)\}$, and there will be a residue in the expansion of $\Phi_l(\lambda; \underline{t})$ on this basis. Hence, for some matrix with $l$, $j$th component $r_{lj}(\underline{t})$,

$$\begin{aligned}\Phi_l(\lambda; \underline{t}) &= \sum_{j=0}^{j=L-1} \alpha_{lj}(\underline{t})\phi_j(\lambda) + F_l(\lambda; \underline{t}) \\ &= \sum_{i=0}^{i=L-1}\sum_{j=0}^{j=L-1} r_{lj}(\underline{t})a_{ji}\phi_i(\lambda) + F_l(\lambda; \underline{t}) \\ &= \sum_{i=0}^{i=L-1} r_{li}(\underline{t})\Phi_i(\lambda; \underline{t}^c) + F_l(\lambda; \underline{t})\end{aligned}$$

where $F_l$ is a residue orthogonal to all the $\phi_m(\lambda)$. Substituting into equation (1), we obtain

$$\begin{aligned}\Psi_n\left(\int \underline{\Phi}(\lambda; \underline{t}^c)s(\lambda)\, d\lambda; \underline{t}\right) &= \sum_{j=0}^{j=L-1} r_{nj}(\underline{t}) \\ &\quad \times \int \Phi_j(\lambda; \underline{t}^c)s(\lambda)\, d\lambda \\ &\quad + \int F_n(\lambda; t)s(\lambda)\, d\lambda \qquad (2)\end{aligned}$$

where $\Psi_n(\cdot; \underline{t})$ is the $n$th component function of $\underline{\Psi}(\cdot; \underline{t})$.

We may expand $s(\lambda)$ in terms of the basis $\{\phi_m(\lambda)\}$, as both $E$ and $S$ are subsets of a space of functions. Thus, there is a *unique* decomposition for any $s(\lambda)$:

$$s(\lambda) = \sum_{i=0}^{i=L-1} \sigma_i\phi_i(\lambda) + s^*(\lambda)$$

where $s^*(\lambda)$ is a residue orthogonal to each of the $\phi_m$. Equation (2) becomes

$$\begin{aligned}\Psi_n\left(\sum_{j=0}^{j=L-1} a_{ij}\sigma_j; \underline{t}\right) &= \sum_{i=0}^{i=L-1}\sum_{j=0}^{j=L-1} r_{ni}(\underline{t})a_{ij}\sigma_j \\ &\quad + \int F_n(\lambda; t)s^*(\lambda)\, d\lambda \qquad (3)\end{aligned}$$

Equation 3 is fundamental to any analysis of color constancy. I call equation (3) the *color constancy equation*, and refer to the term $\int F_n(\lambda; \underline{t})s^*(\lambda)\, d\lambda$ as the residual term. This term is the only impediment to success on the part of a color constancy algorithm.

In effect, the residual term is that part of the receptor response under the illuminant $\underline{t}$ resulting from properties of the surface reflectance function *that were not observed under the canonical illuminant*. Thus, it is not possible to predict the residual term without a constraint on surface reflectance functions, or knowledge of the illuminant. The simplest case occurs when either illuminants or surface reflectances are constrained so that the residual term is always zero. If the residual term satisfies some relation, color constancy may be possible with appropriate constraints on the surface reflectances and illuminants. This case is of little interest, and is not dealt with here, because the constraints required will be very strong. From now on, I assume that constraints on the surface reflectances and the illuminants ensure that the residual term is zero.

Now $\underline{\Psi}(\underline{p}; \underline{t})$ predicts the appearance of a surface under some light represented by $\underline{t}$, given the response $\underline{p}$ from the surface under the canonical light. Color constancy requires estimating $\underline{t}$ and applying $\underline{\Psi}^{-1}(\cdot; \underline{t})$ to the receptor responses observed. Equation (3) shows that constructing some nonlinear $\underline{\Psi}^{-1}(\cdot; \underline{t})$ is erroneous. The residual terms cannot be determined, or accounted for by nonlinearity in the form of this function without other strong assumptions about the form of surface reflectance functions. Thus, color constancy is possible only in a world where these terms are constrained. Since I assume these terms are zero, $\underline{\Psi}^{-1}(\cdot; \underline{t})$ is a linear map.

Three alternative preconditions will cause the residual term to be zero:

1. There is a constraint on the illuminant, such that the residual terms are zero. In particular, for any illuminant, parametrized by $\underline{t}$, the residual term $F_n(\lambda; \underline{t}) = 0$, for all $n$. In a world where this is the case, one can build an algorithm that will work for an infinite dimensional set of surface reflectance

functions if $t$ can be estimated. I show that one can in fact estimate $\underline{t}$ and that this constraint naturally implies a version of the coefficient rule.

2. Surface reflectance functions are only $L$ dimensional, and in particular are members of the space spanned by $\{\phi_0(\lambda), \ldots, \phi_{L-1}(\lambda)\}$. In this case a more general mapping of the receptor responses will achieve color constancy. Constancy will be possible for an infinite dimensional space of illuminants in this case, but only a finite dimensional subset of the illuminants causes any response from the surface at all.
3. In the third case, surface reflectances are $M > L$ dimensional, where $M$ is finite. In this case, we assume that the space of surface reflectances is spanned by

$$\{\phi_0(\lambda), \ldots, \phi_{L-1}(\lambda), \phi_L(\lambda), \ldots, \phi_{M-1}(\lambda)\}$$

where $\phi_0(\lambda), \ldots \phi_{L-1}(\lambda)$ are derived from the receptors and the canonical illuminant in the manner indicated. Now if every possible illuminant is a member of a finite dimensional space of functions such that for every $e_i(\lambda)$ in some basis of the space (and hence for any basis) $\int \rho_k(\lambda)e_i(\lambda)\phi_j(\lambda)\, d\lambda = 0$ for $L \leq j < M$ and $0 \leq k < L$, the residual terms will be zero.

In what follows below, I refer to case 1 as a minimal assumption of the first kind, and to cases 2 and 3 being minimal assumptions of the second kind. This is because the distinction between the case 2 and case 3 assumptions is not strong, whereas the distinction between the two kinds of minimal assumption is large. The minimal assumption of the first kind corresponds to assuming an infinite dimensional set of surface reflectances, and a strong constraint on illuminants described in section 5.1. Minimal assumptions of the second kind correspond to requiring that surface reflectance functions be finite dimensional, with a weaker constraint on the illuminants. These assumptions lead to an algorithm for color constancy that differs in its details depending on the assumptions employed. Since the minimal assumptions of the second kind are stronger than those of the first kind, it is possible to recover more information from an image in a world that satisfies the second kind than in one that satisfies the first kind. Unfortunately, it is difficult to be convinced that minimal assumptions of the second kind are appropriate for real pictures. If none of these assumptions (or stronger ones) apply, color constancy is not possible.

In fact, the results in section 6 suggest that it is sufficient to ensure that residual terms are always small with respect to the receptor responses, at least for minimal assumptions of the first kind.

## 4 Recovering the Illuminant

Assume that in the world in which the color constancy algorithm is going to work, the residual terms are zero. Thus, the illuminants and surface reflectances must be such that under any illuminant, one observes only the degrees of freedom in surface reflectance that one observes under the canonical illuminant. I assume that the world in which a color constancy program works contains only illuminants that have this property. In section 5, I discuss what illuminants this world could contain, but for now assume that the set of such illuminants is nontrivial. Equation 3 becomes

$$\Psi_n\left(\sum_{j=0}^{j=L-1} a_{ij}\sigma_j;\ \underline{t}\right) = \sum_{i=0}^{i=L-1}\sum_{j=0}^{j=L-1} r_{ni}(\underline{t})a_{ij}\sigma_j \qquad (4)$$

This means that the gamut under some illuminant $\underline{t}$ is the image in a linear map of the gamut under the canonical illuminant. As a result, it is possible to use geometrical properties of the gamut to estimate which linear map was applied, and as a result to determine the illuminant.

Intuitively, one observes only a limited number of colors under any given light, because a surface can reflect no more light than falls on it, and cannot reflect less than no light at all. That is, surface reflectance functions are never less than zero, nor greater than one. Thus, they form a bounded, closed, and convex subset of the space of all square integrable functions. Boundedness and closedness are obvious: convexity is true because if $f(\lambda)$ and $g(\lambda)$ are two such functions, $\mu f(\lambda) + (1 - \mu)g(\lambda)$ must lie between zero and one for all $\lambda$ and for $0 \leq \mu \leq 1$, and so this is a third such function. If one were to image every possible surface reflectance under some illuminant, the resulting gamut would also be bounded, closed, and convex, because it is the image in a linear map of this subset.[2] Define the canonical gamut to be the gamut obtained when one images every possible surface reflectance under the canonical illuminant. Assume that the canonical gamut is known.

There is a linear map, $\underline{\Psi}(\cdot; \underline{t})$ corresponding to each illuminant $\underline{t}$, which takes the appearance of a surface

under the canonical illuminant to its appearance under the illuminant $\underline{t}$. By assumption, this map is bijective. Given a picture of a set of surfaces, the gamut of this picture is a subset of the image of the canonical gamut in some bijective linear map. Color constancy requires determining what the map was, and applying the inverse map. Consider these inverse maps. By definition, no observation that originated in a point outside the canonical gamut corresponds to a surface reflectance. We can then immediately rule out any inverse map that takes the observed gamut to a proper superset of the canonical gamut, because this implies that some surface reflectance does not lie in the canonical gamut when imaged under the canonical illuminant.

This process leaves a feasible set of linear maps, any of which might be the inverse of the map associated with the illuminant that formed the image. The minimal assumption made corresponds to a further constraint on this feasible set, because it is effectively a constraint on the form of the illuminant. These constraints are explored below, in section 5. The feasible set is the set of linear maps that satisfy both the gamut constraint and the constraints resulting from the minimal assumptions. The feasible set is thus the set of inverse maps corresponding to the illuminants under which the picture could have been taken. An estimator can be used to choose the most likely inverse map within this set.

Thus, an algorithm for performing color constancy has this form:

1. Construct the canonical gamut, by observing as many surfaces as possible under a single given light. This light will be the canonical illuminant, and surface color is defined to mean the color that the surface has when observed under this light. The canonical gamut can be approximated by taking the convex hull of the union of the gamuts obtained by these observations. Call the canonical gamut $C$.
2. To construct the feasible set for any patch imaged under a constant illuminant:
   - Form the convex hull of the gamut observed. Call this gamut $D$.
   - Form the set of the $L$-by-$L$ matrixes, $M$, of rank $L$ such that $M(D) \subset C$. The particular minimal assumption under which the color constancy algorithm is operating determines a subset of this set, which is the feasible set.
3. Within this feasible set, use some estimator to choose that map most likely to correspond to the illuminant.
4. Apply the chosen map to the receptor responses to obtain color descriptors. The color descriptors are an estimate of the appearance of the surfaces in the original image under the canonical light.

Which minimal assumption is chosen dramatically affects the feasible set. In the next section, I analyze these effects and show how the feasible set can be completely determined.

## 5 The Structure of the Feasible Set

By assumption, the color constancy algorithm encounters only illuminants with particular properties. In this section, I show which illuminants possess these properties, and demonstrate that the restrictions on illuminant properties result in restrictions on the linear maps associated with the illuminants. For minimal assumptions of the first kind, only a small class of linear maps need be considered, and the algorithm can recover few degrees of freedom in the illuminant. For minimal assumptions of the second kind, the number of degrees of freedom in the illuminant that the algorithm can recover depends on geometric properties of the gamut, but is surprisingly large.

### *5.1 Minimal Assumptions of the First Kind*

Assuming that for all $\underline{t}$, $F_k(\lambda, \underline{t}) = 0$ implies, by equation 2, that for some $r_{l_j}(\underline{t})$:

$$\Phi_l(\lambda; \underline{t}) = \sum_{j=0}^{j=L-1} r_{l_j}(\underline{t})\Phi_j(\lambda; \underline{t}^c)$$

Recall that $\underline{\Phi}(\lambda; \underline{t})$ is the vector with components $\Phi_l(\lambda; \underline{t}) = \rho_l(\lambda)e(\lambda; \underline{t})$. Define a vector-valued function, $\underline{\rho}(\lambda)$ such that $\underline{\Phi}(\lambda; \underline{t}) = e(\lambda; \underline{t})\underline{\rho}(\lambda)$. Then for a matrix $R(\underline{t})$ having $l$, $j$th component $r_{l_j}(\underline{t})$, we have

$$\underline{\rho}(\lambda)e(\lambda; \underline{t}) = R(\underline{t})\underline{\rho}(\lambda)e(\lambda; \underline{t}^c)$$

Assume that $R$ has full rank; if it doesn't, one may ignore classes of receptors until the diminished system has full rank, and proceed. If $e(\lambda; \underline{t}) = 0$, then either $e(\lambda; \underline{t}^c) = 0$ or $\underline{\rho}(\lambda) = 0$. This case is of little interest because it means that either the scene is not illuminated, or the surfaces reflect none of the illuminating light. Recall that $e(\lambda; \underline{t})$, and $e(\lambda; \underline{t}^c)$ are scalar-valued functions, and consider

$$\underline{\rho}(\lambda) = R(\underline{t})\,\underline{\rho}(\lambda)\,\frac{e(\lambda;\,\underline{t}^c)}{e(\lambda;\,\underline{t})}$$

Then this requires for all $\lambda$ that $\underline{\rho}(\lambda)$ is an eigenvector of $R(\underline{t})$. But $R(\underline{t})$ is a function of $\underline{t}$ alone and has at most $L$ distinct eigenvectors, so the components of $\underline{\rho}(\lambda)$ must be chosen such that $\rho(\lambda)$ is for any $\lambda$ an eigenvector of $R$. This is possible because $\underline{\rho}(\lambda)$ represents the spectral sensitivity of the sensors, and can be chosen. Since the eigenvectors of $R(\underline{t})$ are fixed by the choice of sensors, the form of $R(\underline{t})$ is restricted. This gives the form of the illuminants for which the algorithm will work.

For $\underline{\rho}(\lambda)$ to be an eigenvector of $R$ for all $\lambda$, there must exist functions $k_j(\lambda)$, such that

$$\underline{\rho}(\lambda) = \sum_{j=0}^{j=L-1} k_j(\lambda)\underline{\epsilon}_j$$

where

$$\text{support}\,(k_i(\lambda)) \cap \text{support}\,(k_j(\lambda)) = \emptyset,\ i \neq j$$

and the $\underline{\epsilon}_j$ are arbitrarily chosen, linearly independent vectors. The terms $\underline{\epsilon}_j$, will be the $L$ eigenvectors of $R(\underline{t})$. These are a result of the choice of receptor. If the eigenvectors $\underline{\epsilon}_j$ are not chosen to be linearly independent, then the receptor sensitivities arising will not be linearly independent.

Given that the eigenvectors of $R$ are fixed, there are $L$ free linear parameters for $R$, that is, its eigenvalues. This means that if minimal assumptions of the first kind apply, the map that achieves constancy consists in form of a change of basis, a diagonal scaling, and another change of basis. In particular, these changes of basis are fixed in advance by the choice of receptors. Furthermore, there exist choices of receptors for which this algorithm cannot work, because they do not have the form given above.

Intuitively, for such a scheme to perform, color constancy requires a strong property of the illuminant. Consider some $\lambda = \lambda_0$; then

$$\underline{\rho}(\lambda_0) = \frac{e(\lambda_0;\,\underline{t}^c)}{e(\lambda_0;\,\underline{t})}\,R(\underline{t})\underline{\rho}(\lambda_0)$$

In particular, as a result of the expansion of $\underline{\rho}$ and the fact that the supports of the coefficient functions are disjoint, there is some $0 \leq i < L$ such that

$$\underline{\rho}(\lambda_0) = k_i(\lambda_0)\underline{\epsilon}_i$$

(no summation over $i$). Then we have

$$\underline{\rho}(\lambda_0) = \frac{e(\lambda_0;\,\underline{t}^c)}{e(\lambda_0;\,\underline{t})}\,R(\underline{t})k_i(\lambda_0)\underline{\epsilon}_i$$

$$k_i(\lambda_0)\underline{\epsilon}_i = \frac{e(\lambda_0;\,\underline{t}^c)}{e(\lambda_0;\,\underline{t})}\,\mu_i k_i(\lambda_0)\underline{\epsilon}_i$$

where $\mu_i$ is the eigenvalue associated with $\underline{\epsilon}_i$. From this equation, one obtains

$$\frac{e(\lambda_0;\,\underline{t}^c)}{e(\lambda_0;\,\underline{t})}\,\mu_i = 1$$

This is true over the support of $k_i(\lambda)$. Thus, minimal assumptions of the first kind require that over the support of each $k_i(\lambda)$, $e(\lambda;\,\underline{t}^c)/e(\lambda;\,\underline{t})$ is constant, where these regions are given by the choice of receptor spectral sensitivity. This is a strong requirement on the illuminant.

Thus, minimal assumptions of the first kind mean that the feasible set consists of all those linear maps that satisfy the gamut restriction and whose eigenvectors are fixed. Since the feasible set is defined, the algorithm is specified. I call this algorithm **Crule**, for the fact that it involves a coefficient rule. The common belief that a coefficient rule is insufficiently general to achieve color constancy originates in the requirement that surface reflectances be finite dimensional. Assuming that surface reflectances belong to a finite dimensional set corresponds to a minimal assumption of the second kind, given that the illuminant functions are also constrained. Minimal assumptions of the second kind are treated in section 5.2.

From the point of view of machine vision, these results are encouraging. If one chooses receptors with disjoint, narrow-band sensitivities, the illuminant will be effectively constant over the support of the receptor. Furthermore, because the sensitivities are disjoint, the receptor sensitivities will be eigenvectors of a diagonal matrix, and the receptor gains can be adjusted independently. Such a system could achieve a very high degree of color constancy for all real surfaces under almost all real lights if the receptors responded to a sufficiently narrow band of wavelengths.

The color constancy equation and the above analysis allows insight into the quality of the results reported by McCann, McKee, and Taylor (1976). Failures from a color constancy system are to be expected as a result of the nature of the problem, and the results that they show for the algorithm they use are extremely good.

Recall, however, that McCann, McKee, and Tay'lor used as an illuminant a weighted sum of narrowband lights, such that each receptor effectively responded to only one light. With this form of illuminant, in all circumstances the map taking receptor responses under one light as compared to those under another such light has only small off-diagonal terms, and that since the lights employed are within a linear map of one another, the residual is always zero (i.e., the minimal assumption of the first kind is always satisfied), so that constancy may be achieved by independent scaling of receptor responses. This experiment does not fully test the Retinex algorithm, because theory shows it should work well in an experiment of this sort. The performance of the algorithm under more general lights remains untested.

Furthermore, McCann, McKee, and Taylor did not note the Retinex algorithms' requirement that some average of surface reflectances remains constant, because they tested it on only one Mondriaan. In section 6, I demonstrate this flaw in the Retinex algorithm, which was originally observed by Brainard and Wandell (1986).

### 5.2 *Minimal Assumptions of the Second Kind*

When either minimal assumption of the second kind is justified, the arguments employed above no longer apply. Intuitively, the fact that this kind of assumption is stronger than the first kind suggests that more parameters in the illuminant can be recovered than could be with the first kind. In fact, using this approach one can recover a surprisingly large number of parameters in the illuminant.

The gamut of an image of all possible surfaces, taken under some unknown illuminant, is the image in a linear map of the canonical gamut. The number of illuminant parameters that can be recovered is the number of degrees of freedom of this linear map that can be computed simply by observing the gamut. This number depends on geometric properties of the gamut itself. For example, consider the set of vectors in $\mathbf{R}^3$ whose length is less than, or equal to, one. By observing its image in a linear map operating on $\mathbf{R}^3$, one may make assertions about the shear components of the map, but not about the rotation components.

For the purposes of the analysis that follows, no distinction need be drawn between the two cases of minimal assumptions of the second kind. Although in the second case the dimension of the space of surface reflectances is higher, the illuminants have been chosen such that only an $L$-dimensional subset of the space of surface reflectances ever causes a receptor response. The second kind of assumption implies that the space of surface reflectances is finite dimensional, and the dimension of the subspace whose effects appear in the receptor responses is at most $L$. If it is less than $L$, some of the receptors are redundant and may be discarded, so that the case where the dimension of this space is precisely $L$ is the only interesting case.

To formalize these intuitions, define $Gs[\Sigma]$ to be the set of bijective linear maps on $\mathbf{R}^L$ that map a subset $\Sigma$ of $\mathbf{R}^L$ to itself. $Gs[\Sigma]$ is clearly a subgroup of $GL(L)$, where $GL(L)$ is the group of bijective linear maps operating on $\mathbf{R}^L$. Call $Gs[\Sigma]$ the *linear similarity* group of the set. Denote by $A(\Sigma)$ the image of $\Sigma$ in the linear map $A$. Then

$$A \in Gs[\Sigma] \Leftrightarrow A(\Sigma) = \Sigma$$

PROPOSITION. For $\Sigma \subset R^L$, and $F, G \in Gl(L, \mathbf{R})$,

$$F(\Sigma) = G(\Sigma) \Leftrightarrow G = FA$$

for some $A \in Gs[\Sigma]$.

*Proof:* (See appendix II.)

PROPOSITION. For $\Sigma \subset \mathbf{R}^L F \in Gl(L)$, $Gs[\Sigma]$ and $Gs[F(\Sigma)]$ are isomorphic.

*Proof:* (See appendix II.)

The important point here is that one cannot tell if members of the linear similarity group of some set have operated on it by observing the image of that set in a linear map. *It cannot "lose linear similarities."*

Now recall that the gamut is the image of the canonical gamut in some bijective linear map. Clearly, for $F \in GL(L)$ and $A$ in $Gs[\Sigma]$, it is impossible to tell the map $F$ from the map $F \circ A$ by observing only the image of $\Sigma$ in these maps. Furthermore, we can distinguish $F$ and $G$ if and only if there is no $A \in Gs[\Sigma]$ such that $F = G \circ A$. Thus, there is a correspondence between the maps we can distinguish and the collection of cosets defined by the right action of $Gs[\Sigma]$ on $GL(L)$. This collection is a manifold only if $Gs[\Sigma]$ is closed in $GL(L)$ (Boothby 1986, p. 94). Assume that this is the case, as the most likely candidates for $Gs[\Sigma]$ are either a finite discrete group of rotations, or a one-parameter rotation group. Then the collection of cosets can be written as $GL(L)/Gs[\Sigma]$.

The set of maps taking the canonical gamut to a superset of the observed gamut is a subset of $GL(L)$. Call this set $\mathfrak{F}$, and write $\mathfrak{C}$ for the canonical gamut. $\mathfrak{F}$ needs to be collapsed to represent the distinctions in the illuminant that can actually be drawn. The wisest strategy is in constructing the system to determine one representative of each right coset of $Gs[\mathfrak{C}]$ in $GL(L)$ corresponding to a legitimate illuminant (call this a *coset leader*). Then the set of matrixes corresponding to the *observable* constraint set is obtained by replacing each member of $\mathfrak{F}$ by its coset leader. This is likely to be computationally laborious, but can be simplified by a sensible choice of coset leaders.

This discussion, in conjunction with section 4, gives a complete specification for a color constancy algorithm, since the set of linear maps that can be distinguished is determined. I call this algorithm **Mwext**, because it is the natural extension of Maloney and Wandell's work, with physical realizability incorporated. However, the way that this algorithm has been defined means that it can recover more parameters in the illuminant than did Maloney and Wandell, for the same number of receptors.

Now the number of illuminant parameters that can be recovered is exactly the dimension $GL(L)/Gs[\mathfrak{C}]$, which may be rather large. Theorem 9.2 of (Boothby 1986, p. 166) states that for Lie groups (the groups involved are Lie groups) $G$ and $H$, $H$ a closed subgroup of $G$, $G/H$ has a unique structure as a differentiable manifold, and the dimension of $G/H$ is dim $(G)$ − dim $(H)$. Thus, if the linear similarity group is a discrete group, the dimension of the manifold of observably distinct illuminants is $L^2$, and if the linear similarity group is a one-parameter subgroup, the dimension is $L^2 - 1$. This is an embarrassment of parameters for a model of the human vision system. People do not recover either 9 or 8 parameters in the illuminant.

This formalism, apart from determining exactly how many parameters are distinguishable in the illuminant, clarifies aspects of Maloney and Wandell's algorithm (Maloney and Wandell 1986; Maloney 1986; Wandell 1987), of which it is in some sense a natural generalization. Maloney and Wandell (1986) formulated a color constancy technique using finite dimensional linear models of illuminant and surface reflectance. The mathematical details of this algorithm are in Maloney's thesis, (Maloney 1984). The description given here is somewhat more abstract than the one Maloney presents, but has the advantage of compactness.

Maloney and Wandell represent surface reflectance functions by $\mathbf{R}^N$, and have $M$ receptors, $M > N$. They use a linear sensor and require that no two unequal surface reflectances produce the same receptor response under any light. Then taking an image corresponds to applying an injective linear map $F$, $F : \mathbf{R}^N \to \mathbf{R}^M$. In particular, if $F(S)$ denotes the image of the set $S$ in the map $F$, the gamut observed is $F(\mathbf{R}^N)$. It is possible to distinguish between maps $F$ that take $\mathbf{R}^N$ to distinct $N$-dimensional subspaces through the origin in $\mathbf{R}^M$. Maloney (1986) discusses the case where there are two degrees of freedom in surface reflectance, and three receptors. Under these conditions, he can recover either two degrees of freedom in surface reflectance and one in the illuminant or one in surface reflectance and two in the illuminant. The asymmetry is due to a scaling ambiguity between surface lightness and illuminant brightness.

I shall discuss this particular example for simplicity, although with little difficulty both the number of receptors and the number of degrees of freedom in the surface reflectance functions may be changed, as long as there are more receptor bases than surface reflectance bases. Recall that they do not require physical plausibility of surface reflectances, so the set of representations of surface reflectance functions becomes a space, $\Sigma_w = \mathbf{R}^2 \times (0)$.

The linear similarity group of a two dimensional linear subspace of $R^3$ is the group of matrixes $M$ which, in some basis, have the form

$$\begin{matrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 0 & \cdot \end{matrix}$$

such that det $(M) \neq 0$. Again, this is easily seen to be a group which, using the notation above, is denoted $Gs[\Sigma_w]$. In particular, the dimension of this (Lie) subgroup considered as a differentiable manifold is 7 (count the dots!). Thus, it is possible to recognize as distinct objects only maps corresponding to distinct points on $GL(3)/Gs[\Sigma_w]$ (this is an object known as a Grassmannian (Boothby 1986), felicitously named after Hermann Grassmann, a pioneer in color vision). To recover a representation of surface reflectance, therefore, there must be only one illuminant corresponding each point in $GL(3)/Gs[\Sigma_w]$. The dimension of $GL(3)$ is 9, and so it is possible to distinguish $9 - 7 = 2$ dimensional manifold of illuminants. Hence, Maloney and Wandell's approach requires a priori constraint on *both* illuminant and surface reflectance.

**Mwext** can recover more parameters in surface reflectance than Maloney and Wandell did, for the same number of receptors. Maloney and Wandell point out that their approach is incapable of uniquely determining surface lightness. This is as a result of the linear similarity:

$$\begin{matrix} k & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & k \end{matrix}$$

(where $k \neq 0$) of the plane. Since the canonical gamut does not have this similarity, it is possible for **Mwext** to recover surface lightness. This, however, depends on the assumption that there are many different colors in the image. If the image contains only very dark colors, for example, the brightness of the illuminant will be poorly constrained, and **Crule** will report surface lightness incorrectly.

## 6 Experimental Results

Very few experimental demonstrations of color constancy programs have been described in the literature. Published tests exist for Maloney and Wandell's (1986) algorithm working on synthetic images only (Maloney 1984). With the exception of results published on one Mondriaan image by McCann, McKee, and Taylor (1976), no published account of the color constancy performance of the Retinex algorithm on real data exists. Brainard and Wandell (1986) tested their model of the Retinex algorithm on synthetic data. Gershon's algorithm (Gershon 1988) has been demonstrated on a single real image. This result is flawed by the fact that there was only a single object on a black background in the image. Brill (1979) implies that he had a color constancy algorithm that worked on real pictures, but I have been unable to find details of this work elsewhere in the literature. Buchsbaum's paper (Buchsbaum 1980) does not disclose whether the results he presents originate in real or in synthetic images.

### *6.1 Preliminary Information*

All images were taken with a monochrome CCD camera with its gain control defeated. Three separate exposures of each image were made using Kodak Wratten filters (no.'s 29, 47B, and 58) for color separation, and a sharp near infrared cut filter, which is essential as a result of the pronounced near infrared sensitivity of CCD cameras. The lenses used were conventional photographic lenses. The CCD camera appears to measure no chromatic aberration, probably because its pixels are large with respect to any fringes. The illuminants used were two 500W photographer's lamps of unknown color temperature, and a warm white appearance. The illuminants were colored by the use of translucent colored plastic filters hung in front of the bulbs. At no stage have the properties of these filters been measured. I refer to the light provided by these lamps as "white light," and to colored illuminants by the conventional color names of the plastic filters used.

Neutral density filters were used to weight the separation filters so that the aperture of the lens did not need to be adjusted between exposures. I chose the weights so that images closely resembled the objects imaged when displayed on the screen of a Sun workstation. Choosing weights on the basis of the overall appearance of the image appears to distribute the errors in the color displayed evenly, making it easier to compare objects with images. A number of other strategies for choosing weights are possible. For example, one might choose the set of weights that led to a bright white patch causing the same response in all three separations, although monitor nonlinearities can cause this approach to skew the color of darker patches.

In fact, the choice of weights is not particularly important in demonstrating color constancy. A color constant algorithm should produce near constant, nontrivial, color descriptors for objects, when presented with well-populated scenes imaged under widely varying lights. It is possible to tell whether an algorithm can do this without ever knowing how the color descriptors produced by the algorithm relate to surface color as people see it. Dealing with an algorithm on such abstract terms is unattractive, however, and a choice of weights that makes the image look recognizable avoids this problem. Color constancy, or any lack of it, is easily recognized from the descriptors.

For this series of experiments, I used Mondriaans made of Color-aid papers (a set of 202 papers, with standard colors, available from the Color Aid corporation[3]) and recorded the colorimetric description (supplied by the manufacturers of the paper, in terms of their own color space) of each patch for comparison and calibration. The patches that comprised each test Mondriaan were chosen from a shaken bag, to provide some

randomness in the structure of their gamut. Each Mondriaan contained 60 patches of paper. Figures 1–3 show the gamuts of images of these Mondriaans taken under different colored lights, and confirm the claim that the gamut is skewed by colored lights.

### 6.2 Implementation Details

**Crule** is simple to implement. Since the support of the photoreceptor spectral sensitivities is nearly disjoint, the matrixes in the feasible set can be approximated with diagonal matrixes.

A canonical gamut can only be approximated. The gamut of any set of observations of a finite set of surfaces consists of a cloud of points. Since we know that the canonical gamut is convex, we can approximate it by the convex hull of this cloud of points. I refer to this polyhedron as the observed canonical gamut. The observed canonical gamut was formed by imaging 180 of the set of 202 Color-aid papers under white light.

Although the gamut that we observe in an image is just a cloud of points, every point in the convex hull of this gamut corresponds to a real surface reflectance. This is because the canonical gamut is convex, and the observed gamut is just a sampling of the image of the canonical gamut in a linear map. As a result, we can consider the convex hull of this gamut. I shall call this polyhedron the observed gamut.

The feasible set consists of those diagonal maps that take the observed gamut to a subset of the observed canonical gamut. The observed gamut is convex, the observed canonical gamut is convex, and the maps are linear. As a result, any map that takes every vertex of the observed gamut to a point inside the canonical gamut, takes the observed gamut to a subset of the canonical gamut, and therefore lies in the feasible set.

We wish to construct the set of diagonal maps that take every vertex of the observed gamut to a point inside the observed canonical gamut. To do this, for each vertex of the observed gamut, we form the set of maps that take this vertex to a point inside the canonical gamut, and intersect these sets. In turn, the set of maps that take a single vertex $p$ of the observed gamut to a point inside the observed canonical gamut is a convex polyhedron, which I shall refer to as $\mathfrak{M}_p$. For a proof that $\mathfrak{M}_p$ is a convex polyhedron, see appendix I.

The vertices of $\mathfrak{M}_p$ consist of those maps that take a vertex, $p$, of the observed gamut, to a vertex of the observed canonical gamut (proved in appendix I). Because the maps are diagonal, the vertices of $\mathfrak{M}_p$ (and hence $\mathfrak{M}_p$ itself) are easily computed for any $p$. The final implementation works as follows:

- The convex hull of the observed gamut is computed.
- For each vertex $p$ of this hull, $\mathfrak{M}_p$ is computed.
- Intersect all the sets computed in this way. This process is relatively simple, because the sets are convex polyhedra in three dimensions. The result is the feasible set.
- Within the feasible set, apply an estimator to choose the map most likely to achieve color constancy. Apply this map to the image, to obtain the color descriptors.

This algorithm is relatively simple to implement, but intersecting the convex hulls requires care as 3D convex hulls are rather difficult to manipulate. Explicitly intersecting the hulls is an unwise technique to use, as it tends to generate clouds of hull points nearly on the same plane, with attendant difficulties of representation and computation. The problem is better approached using cuboid approximations (see Cameron 1989; Woodwark and Quinlan 1984).

The estimator used simply chose the map that gave the gamut with the largest volume. This means that the image of the mapping of the gamut of the original picture will fit inside the gamut obtained under the canonical illuminant (because we have chosen a map in the constraint set), and will be the "largest" such image. The parameters corresponding to this map are easily found by noting that a diagonal linear map takes a volume to that volume multiplied by the trace of the map. Thus, the feasible set is simply searched for the map with the largest trace.

This discussion implicitly assumes that the observed gamut is a good approximation of the gamut that would have been seen if every possible surface reflectance were shown under the illuminant. If the observed gamut does have this property, then the feasible set will be small. If it doesn't, the feasible set can be large, and the estimator is more likely to err. This case occurs when the surface reflectances in the scene are not well distributed—for example, when the scene contains only shades of one or two colors.

### 6.3 Results

**Crule**'s performance has been tested on three Mondriaans each of sixty chips of colored paper, each viewed under six different lights. The Brainard and Wandell

model of the Retinex algorithm (Brainard and Wandell 1986) was implemented for comparison. Both algorithms achieve roughly the same degree of constancy, with the Retinex algorithm perhaps slightly outperforming the **Crule** algorithm, when tested on a single Mondriaan viewed under many different lights. Color figures 1 through 6 show the inputs for each algorithm. Although the images appear widely skewed in color, during the imaging it was possible for human observers to give reasonable color names to the Mondriaan patches. Color figures 7 through 12 show the outputs from **Crule** for the given images. The images are similar in color, suggesting that **Crule** is capable of good color constancy. However, the output for an image taken under red light, shown in color figure 8, shows that **Crule** performed poorly on this image. In color figure 9, the output of **Crule** for red light has been omitted because the algorithm failed as badly. This is a quantization effect, discussed more fully in section 7.

Notice that **Crule** has performed rather more poorly on the third Mondriaan than on the other two. This is because the third Mondriaan has a narrower range of colors in it, so that the illuminant is less well constrained.

Figures 4 through 6 show the receptor responses observed for a set of color chips, selected from the first Mondriaan to give a fair impression, under six different lights. These are widely scattered, demonstrating how strongly illuminant color can skew receptor responses. Figures 7 through 9 show the descriptors computed by **Crule** for the same set of chips, from images taken under six different lights. The descriptors are far less widely scattered, indicating that **Crule** is capable of good color constancy.

Color figures 13 through 18 show the outputs of the Retinex algorithm, on the same images, for comparison. The Retinex algorithm is also susceptible to the quantization effect mentioned above, as color figures 13 and 15 show. On the whole, for the case of a single Mondriaan and a set of different lights, the Retinex algorithm appears to out perform **Crule**. However, the Retinex algorithm depends on a geometric average of surface color remaining constant, and is confounded when, for example, red borders are appended to a Mondriaan to change this average (see color figure 19). Large failures in constancy can be caused in this way, as color figure 20 demonstrates. This figure compares the performance of **Crule** and the Retinex algorithm on an image of the first Mondriaan to their performance when red borders have been attached to the Mondriaan, both viewed under white light. Because the Retinex algorithm is sensitive to the spatial average of surface color, a change in this average skews its surface color descriptors. This causes the slight green cast to the output of the Retinex

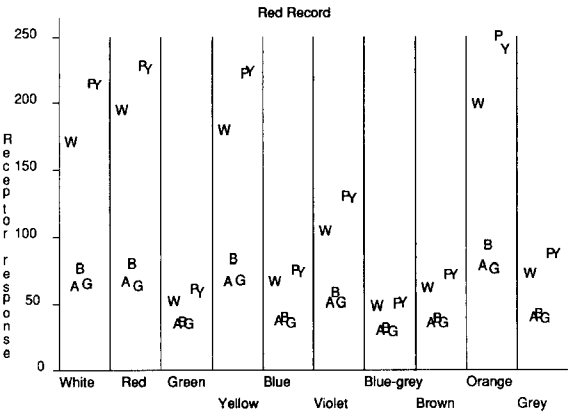



*Fig. 4.* This and the next five figures demonstrate the performance of **Crule** for 10 chips, selected to give a fair impression. This and the next two figures show the receptor responses for 10 chips, under six different lights. Each column represents a chip and is labeled with a color name for each chip under incandescent light. This figure shows the responses for the camera's red channel. The response under white light is plotted as a "W," under blue-green light as an "A," under blue light as a "B," under green light as a "G," under purple light as a "P," and under yellow light as a "Y." See section 6.1 for the meaning of the color names for the illuminants. Note the wide spread of responses, which will make it difficult to use these values to describe chips. The precise color name, in the Color Aid corporation's scheme, for each chip is as follows. The chip labeled white is white, red is RO-Hue, yellow is Y-T1, blue is BGB-S3, violet is VRV-S1, blue-grey is GYG-S1, brown is OYO-S1, orange is YO-T3, and grey is grey.

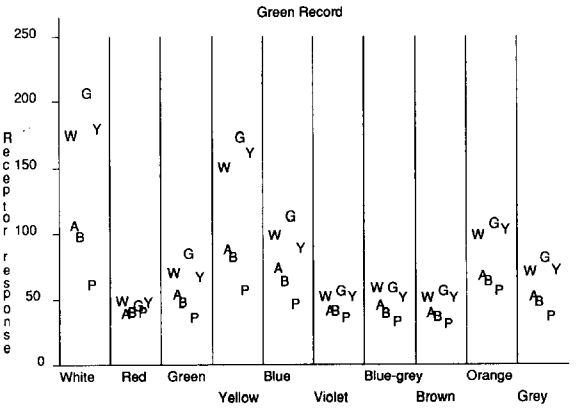



*Fig. 5.* This figures shows the receptor responses for the camera's green channel, using the same conventions as figure 4.

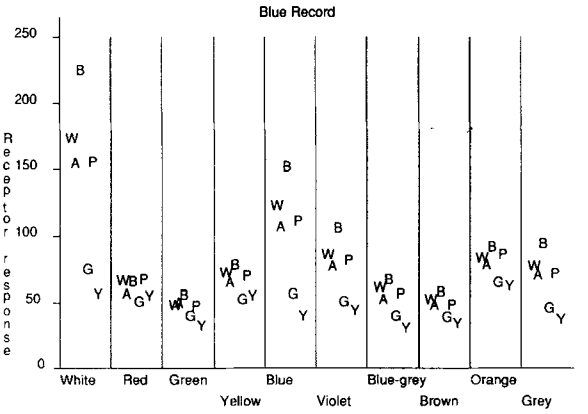



*Fig. 6.* This figure shows the receptor responses for the camera's blue channel, using the same conventions as figure 4.

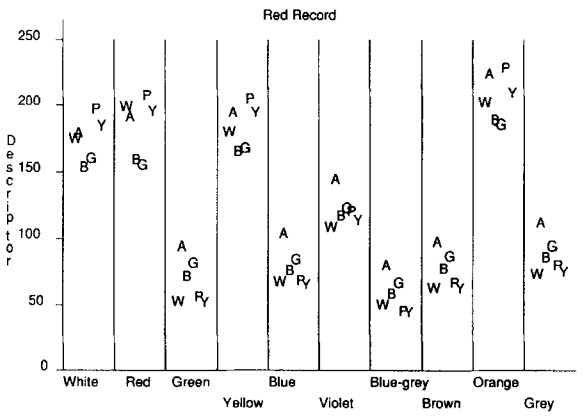



*Fig. 7.* This figures shows the descriptors in the red channel, output by **Crule**, using the same conventions as figure 4. Notice that the descriptors are clustered, indicating a high degree of color constancy.

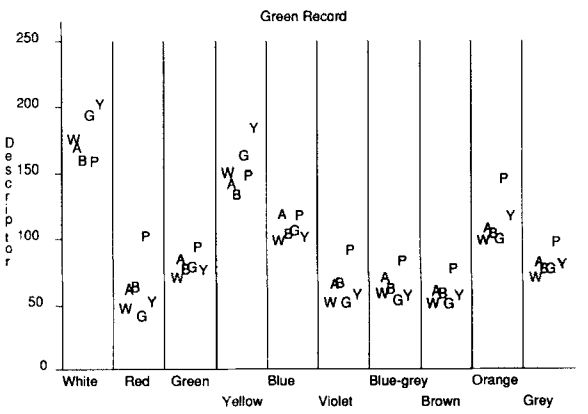



*Fig. 8.* This figure shows the descriptors in the green channel, output by **Crule**, using the same conventions as figure 4. Again the descriptors are clustered, indicating a high degree of color constancy.

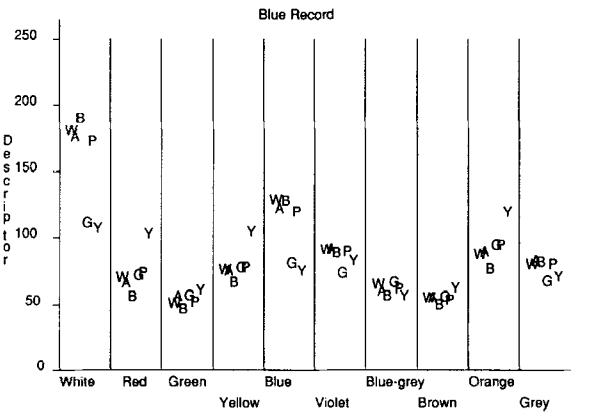



*Fig. 9.* This figure shows the descriptors in the blue channel, output by **Crule**, using the same conventions as figure 4. Again, the descriptors are clustered, indicating a high degree of color constancy.

for the Mondriaan with red borders in color figure 20. This effect can be made very large by increasing the size of the borders, or by comparing images with different colored borders. Since **Crule** does not depend on the spatial extent of a color, the borders do not affect its descriptors.

A statistical analysis is desirable, because images may conceal misbehavior on the part of the algorithms. There is no standard measure of color constancy. The analysis shown uses the median Euclidean distance of the outputs from the average (over the different lights) output, for each chip. This median is then normalized by the Euclidean magnitude of the outputs. Clearly, this statistic measures clustering, and will be zero for every chip for a perfect color constancy algorithm; and as the algorithm delivers increasingly poor performance, the statistic will increase. Normalization is necessary to prevent an algorithm improving its apparent performance by multiplying its outputs by a small number.

The graphs in figures 10 through 12 show the cumulative distribution for this statistic, for two separate images of sixty chips under six different lights (white, green, blue-green, yellow, and purple). The results for both **Crule** and Retinex under red light were skewed for reasons described in section 7, and were omitted from this analysis. The graphs plot the statistic for descriptors computed from images of the first Mondriaan, and the statistic for descriptors computed for images of the first Mondriaan with and without red borders. The red borders skew the descriptors computed by the Retinex algorithm, and the second plot for the Retinex algorithm reflects this substantially larger scatter in its descriptors. This can

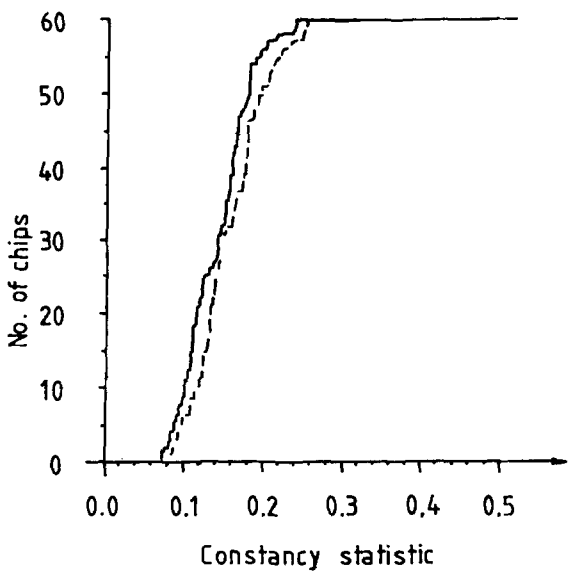



*Fig. 10.* Cumulative distribution of the statistic described in the text for sixty descriptors computed by **Crule**, operating on images of the first Mondriaan without borders (plotted as a solid line), and on images of the first Mondriaan both with and without borders (plotted as a dashed line). The statistic measures the scatter of the descriptors computed for the chips in images under different lights—the larger the statistic, the wider the scatter, and the poorer the algorithm. Note that adding colored borders to the Mondriaan imaged does not significantly affect the descriptors that **Crule** computes, as expected.

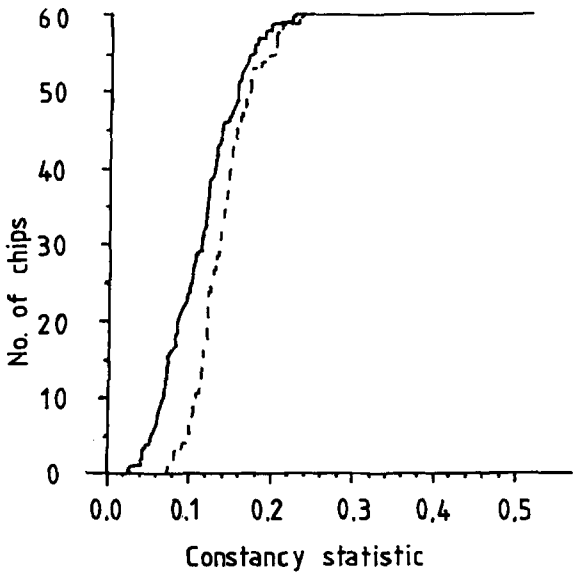



*Fig. 11.* Cumulative histogram of the statistic described in the text for the outputs of the Retinex algorithm operating on images of the first Mondriaan without borders (plotted as a solid line), and on images of the first Mondriaan both with and without borders (plotted as a dashed line). The statistic measures the scatter of the descriptors computed for the chips in images under different lights—the larger the statistic, the wider the scatter, and the poorer the algorithm. Adding colored borders to the Mondriaan imaged affects the descriptors computed, as expected. Using larger borders would increase the scatter of the descriptors. Adding green borders to one image, and red ones to another, would increase it further still. The cause of the effect is described in the text.

be made very large, by appending large borders, or by testing the Retinex algorithm on one scene with green borders, one with red borders, and one without borders. This effect is a basic property of any algorithm that assumes that some spatio-temporal average of surface color is constant. Other algorithms with this flaw include those of Buchsbaum (1981) and Gershon (1988). This skewing effect can happen easily in real images: for example, one does not expect the average surface color of a view of a forest in spring to be the same as that of the same view, taken in autumn.

**Crule** does not have this flaw, but if one considers descriptors for only one image under many lights, the Retinex algorithm performs better than **Crule**, because in this case the average surface color is trivially constant. This may be alleviated by computing the feasible set to higher accuracy; at present, the descriptors computed by the new algorithm may have errors as large as 2% of full range simply as a result of the scheme adopted for computing 3D convex hulls, which makes computing the feasible set to high accuracy expensive in space.

## 7 Discussion

**Crule** is simple to implement, and works well for real pictures. **Crule** is successful because the constraint on surface reflectance is true for all surfaces, and strongly constrains the colors possible in an image.[4] *Representing* surface color using a feasible set has considerable attraction. This representation follows easily from a feasible set representation of the illuminant. Simply, a particular pixel can have arisen only under certain lights, and corresponding to each feasible light is the surface color that will generate that pixel value under that light. Thus, there is a feasible set of surface colors corresponding to each pixel. Given that a feasible set of surface colors can be achieved at a very early stage of visual processing, it seems reasonable to maintain this representation until the last possible moment. If we do this, other sources of visual information may provide cues that reduce the size of the feasible set—shape and surface

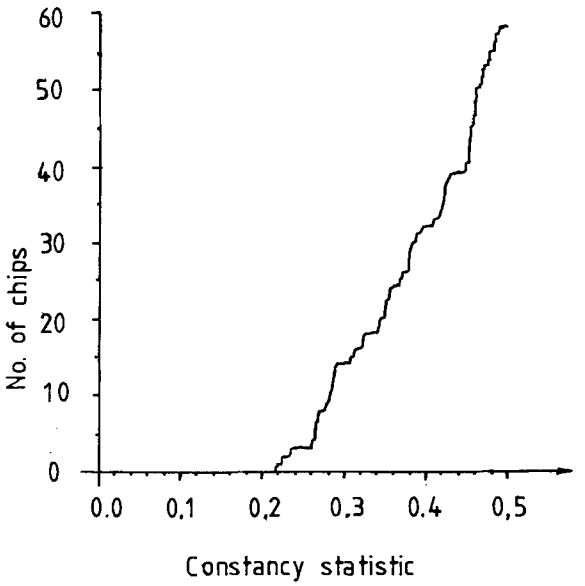



*Fig. 12.* Cumulative histogram of the statistic described in the text for the first Mondriaan. The statistic measures the scatter of the descriptors for the chips in images under different lights—the larger the statistic, the wider the scatter. In this case, the statistic is computed for the receptor responses, so that one can confirm that both **Crule** and the Retinex algorithm can produce descriptors that are an improvement on the receptor responses alone.

color, for example, are deeply entwined by mutual illumination effects (Gershon et al. 1986; Forsyth and Zisserman 1989; Forsyth and Zisserman 1990; Ho et al. 1989).

**Crule** will produce color descriptors for two color images, of the kind used by, for example Land (Land 1959a; Land 1959b). In these experiments Land took slides of a series of scenes. One slide was exposed only in the red band, the other only in the green band, so that for each scene he had a red slide and a green slide. The red separation was inserted in a slide projector casting red light, and the green separation in a projector casting white light. The two slide images were shown in register to an audience, which reported a wide range of colors. Naive colorimetric methods would predict that the images would contain only a range of shades of pink. This effect was not in fact first noticed by Land; the history of this effect is well presented in (Judd 1960).

It would be a definite advantage if **Crule** was able to reproduce this property. A simple geometric argument suggests that this is the case. The gamut of a two-color image is on a triangular planar patch in surface receptor space, with one corner at the origin, one corner at the "brightest white" point, and one corner at a "deep red" point. We can reasonably expect to be able to construct a map that takes this into a larger patch within the gamut.

Two color images were constructed from conventional color imges, by mapping the green and blue components to $r$ times the original green component, and the red component to the sum of $r$ times the original green and $s$ times the original red components, where $0 \leq r$, $s \leq 1$. These receptor responses model the receptor responses expected from the slide experiment.

**Crule** has been run on such images. The result is shown in color figure 21. No claim is made that this result is quantitatively similar to human performance, only that the algorithm can infer a broad gamut from a two-color image. This is obviously the case, although the images display in neither the original data nor in the photographs quite the breadth of gamut that Land describes (he talks of "lovely color photographs" (Land 1959a, p. 120). Green patches are conspicuous by their absence, as are yellow patches.

Both **Crule** (color figure 8) and the Retinex algorithm (color figures 13 and 15) performed very poorly on images taken under red light. This failure is explained as follows. There is very little information in the input data either in the green or in the blue channels, because to keep the camera from saturating in the red channel, the aperture had to be closed. It would clearly be unfair to attempt to demonstrate the effectiveness of an algorithm for color constancy that employs the coefficient rule, and then to adjust the aperture separately for each color exposure, for that would presuppose a constancy algorithm. Equally, if this is not done, *constancy is impaired* by measurement errors.

The constancy algorithm is connected to its receptors by a quantizing system with constant quantization thresholds. As a result, under red light, the signal on the green channel has a dynamic range of about three bits. The constancy algorithm is thus more ignorant of what the green transducer is measuring, than it is about what the red transducer is measuring, and its constancy is impaired. The human vision system faces a similar problem, in that the operating range of human photoreceptors is very much greater than the signaling capacity of the nerve channels (Barlow and Mollon 1982). In humans, this problem is accounted for by a process known as adaptation, which is described in more detail elsewhere (for example, Bartleson 1977; Beck 1972; Wyszecki and Stiles 1982). The effect of adaptation is

to use a spatio-temporal average of the receptor's inputs to adjust its gain so that it can measure effects over a wide dynamic range. The color constancy of a machine vision system could be significantly improved by a similar approach. One might, for example, use a low bandwidth control loop to ensure that the spatio-temporal average of the video signal is constant, by adjusting the integrating time of the CCD camera. By supplying the integrating time information to the color constancy system, one could improve its performance on deeply colored lights.

This work had not addressed the case where the illuminant varies over space. The Retinex algorithm, and algorithms that devolve from it (Horn 1974; Blake 1985; Brelstaff and Blake 1987; Funt and Drew 1988) represent the only existing implemented techniques for dealing with this problem. These algorithms work by declaring a spatial frequency threshold, and asserting that changes below this threshold are due to lighting effects, and changes above it arise from surface color effects. It is unclear how well such a technique will work for other than Mondriaan images. There is every reason to believe that illuminant effects occur at high spatial frequencies (Gilchrist 1983). This suggests that complex processes are necessary to determine whether a change in an image occurred as a result of a change in surface color or a change in lighting, before color constancy programs can work successfully in any but very constrained worlds.

## 8 Conclusion

By analyzing the circumstances under which color constancy is possible, I have developed a color constancy algorithm, **Crule**, and have demonstrated that it achieves color constancy on real images of Mondriaan's. **Crule** estimates the illuminant in colored pictures using a system of constraints. These contraints derive from physical restrictions on the form of surface reflectance functions.

The performance of **Crule** on real images compares favorably with that of the Retinex algorithm of Land (Land and McCann 1971) when run on single Mondriaan images. However, changing the spatial average of surface reflectance disrupts the color constancy of the Retinex algorithm, but does not affect **Crule**. I have shown that **Crule** can mimic qualitatively Land's (Land 1959a; Land 1959b) work on two-color images. This experimental work lead us to see chromatic adaptation as a measurement strategy employed by humans, and to realize that machines require a similar strategy to perform accurate constancy.

Although this algorithm gives very good results, its success at computing surface color depends very much on the unrealistic assumptions that underly the Mondriaan world. These may fail in the real world: for example, illuminants change quickly spatially (Gilchrist 1983). It is clear that surface color algorithms in vision systems will need to be able to cope with more interesting objects than flat, frontally presented surfaces. There is much scope for further work.

## Acknowledgments

I thank Michael Brady, who supervised this work, for his help and support. Margaret Fleck discussed this work at length with me. Andrew Blake, Don Pearson, and Tom Troscianko have made helpful comments. Brian Wandell and Laurence Maloney patiently helped me to understand how and why their techniques worked. An anonymous referee made numerous, careful comments that were an important help in making this paper readable.

## Notes

1. This assumes that there is always a wide variety of surface reflectances in the scene.
2. Koenderink (1987) observes that Luther (1921) and Nyberg (1928) were aware of this result.
3. Address: 37 East 18th St., New York, NY 10003.
4. This constraint has been successfully exploited by Kawata et al. (1987) in a system that segments microscope images, based on dye concentrations.

## References

H.B. Barlow and J.D. Mollon, *The Senses*. Cambridge University Press: Cambridge, 1982.

C.J. Bartleson, "A review of chromatic adaptation." In F.W. Billmeyer and G. Wyszecki (eds.), *Color 77*. Adam Hilger: Bristol, 1977.

J. Beck, *Surface Color Perception*. Cornell University Press: Ithaca, NY, 1972.

A. Blake, "Boundary conditions for lightness computation in Mondriaan world." In D. Ottoson and S. Zeki (eds.), *Central and Peripheral Mechanisms of Color Vision*. Macmillan: New York, 1985.

W. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry.* Academic Press: San Diego, CA, 1986.

G. Buchsbaum, "A spatial processor model for object color perception," *J. Franklin Inst.* 310: 1–26, 1980.

D.H. Brainard and B.A. Wandell, "Analysis of the Retinex theory of color vision," *J. Opt. Soc. Amer. A* 3: 1651–1661, 1986.

G.J. Brelstaff, "Inferring surface shape from specular reflections." Ph.D. thesis, University of Edinburgh: Edinburgh, 1988.

G. Brelstaff and A. Blake, "Computing lightness," *Pattern Recognition Letters* 5: 129–138, 1987.

M.H. Brill, "A device performing illuminant-invariant assessment of chromatic relations," *J. Theor. Biol.* 71: 473–478, 1978.

M.H. Brill, "Computer simulation of object color recognisers," *J. Opt. Soc. Amer.* 69: 1405, 1979.

M.H. Brill and G. West, "Contributions to the theory of invariance of color under the condition of varying illumination," *J. Math. Biol.* 11: 337–350, 1981.

S.A. Cameron, "Efficient intersection tests for objects defined constructively," *Intern. J. Robotics Res.* 8: 1, 1989.

M. D'Zmura and P. Lennie, "Mechanisms of color constancy," *J. Opt. Soc. Amer. A* 3 (10): 1662–1672, 1986.

R.K. Eisenschitz, *Matrix Algebra for Physicists.* Heinemann: London, 1966.

D.A. Forsyth, "A novel algorithm for colour constancy," *Proc. 2nd Intern. Conf. Comput. Vision*, Tampa, 1988a.

D.A. Forsyth, "Colour constancy and its applications in machine vision." Ph.D. thesis, University of Oxford, Oxford, 1988b.

D.A. Forsyth and A. Zisserman, "Mutual illumination," *Proc. 26th IEEE Conf. Comput. Vision and Pattern Recog.*, Las Vegas, 1989.

D.A. Forsyth and A. Zisserman, "Shape from shading in the light of mutual illumination," *Image and Vision Computing*, 1990 (in press).

B.V. Funt and M. Drew, "Color constancy computation in near-Mondriaan scenes using a finite dimensional linear model," *Proc. IEEE Conf. Comput. Vision and Pattern Recog.* 1988.

R. Gershon, "The use of colour in computational vision." Ph.D. thesis, University of Toronto, 1988.

R. Gershon, A. Jepson, and J. Tsotsos, "Ambient illumination and the determination of material changes," *J. Opt. Soc. Amer. A* 3: 1700–1707, 1986.

A.L. Gilchrist, S. Delman, and A. Jacobsen, "The classification and integration of edges as critical to the perception of reflectance and illumination," *Perception and Psychophysics* 33: 425–436, 1983.

J. Ho, B.V. Funt, and M.S. Drew, "Disambiguation of illumination and surface reflectance from spectral power distribution of color signal: Theory and applications," CSS/LCCR TR 88-18, Centre for Systems Science, Simon Fraser University, 1988.

J. Ho, M.S. Drew, and B.V. Funt, "Color constancy from mutual reflection," CSS/LCCR TR 89-02, Centre for Systems Science, Simon Fraser University, 1989.

B.K.P.H. Horn, "Determining lightness from an image," *Comput. Graph. Image Proc.* 3: 277–299, 1974.

D.B. Judd, "Appraisal of Land's work on two-primary colour projections," *J. Opt. Soc. Amer.* 50: 254–268, 1960.

S. Kawata, S. Sasaki, and S. Minami, "Component analysis of spatial and spectral patterns in multispectral images. I. Basis," *J. Opt. Soc. Amer. A* 4: 2101–2106, 1987.

G.J. Klinker, S.A. Shafer, and T. Kanade, "Using a colour reflection model to separate highlights from object colour," *Proc. 1st Intern. Conf. Comput. Vision*, London, 1987.

J.J. Koenderink, "Color atlas theory," *J. Opt. Soc. Amer. A* 4: 1314–1321, 1987.

J. von Kries, "Beitrag zur Physiologie der Gesichtsempfindung," *Arch. Anat. Physiol.* 2: 505–524, 1878.

E.H. Land, "Color vision and the natural image. Part I," *Proc. Natl. Acad. Sci. U.S.A.* 45: 115–129, 1959a.

E.H. Land, "Color vision and the natural image. Part II," *Proc. Natl. Acad. Sci. U.S.A.* 45: 636–644, 1959b.

E.H. Land, "Color vision and the natural image. Part III: Recent advances in Retinex theory and some implications for cortical computations," *Proc. Natl. Acad. Sci. U.S.A.* 80: 5163–5169, 1983.

E.H. Land, "Recent advances in Retinex theory." In D. Ottoson and S. Zeki (eds.), *Central and Peripheral Mechanisms of Colour Vision.* Macmillan: New York, 1985.

E.H. Land, "Recent advances in Retinex theory," *Vision Research* 26: 7–21, 1986.

E.H. Land and J.J. McCann, "Lightness and Retinex theory," *J. Opt. Soc. Amer.* 61 (1): 1–11, 1971.

H.-C. Lee, "Method for computing the scene-illuminant chromaticity from specular highlights," *J. Opt. Soc. Amer. A* 3: 1694–1699, 1986.

R. Luther, "Aus dem Gebiet der Fabreizmetrik," *Z. Techn. Phys.* 12: 540–558, 1921.

L.T. Maloney, "Computational approaches to color constancy." Ph.D. dissertation, Stanford University, Stanford, CA, 1984.

L.T. Maloney, "Evaluation of linear models of surface spectral reflectance with small numbers of parameters," *J. Opt. Soc. Amer. A* 3 (10): 1673–1683, 1986.

L.T. Maloney and B.A. Wandell, "A computational model of color constancy," *J. Opt. Soc. Amer. A*1: 29–33, 1986.

J.J. McCann, S.P. McKee, and Taylor, "Quantitative studies in Retinex theory," *Vision Research* 16: 445–458, 1976.

N. Nyberg, "Zum Aufbau des Farbkoerpers im Raume aller Lichtempfindungen," *Z. Phys.* 52: 506–419, 1928.

P. Sallstrom, "Colour and physics: some remarks concerning the physical aspects of human colour vision," University of Stockholm: Institute of Physics Report 73-09, 1973.

W.S. Stiles, G. Wyszecki, and N. Ohta, "Counting metameric object-color stimuli using frequency limited spectral reflectance functions," *J. Opt. Soc. Amer.* 67 (6): 779–784, 1977.

B.A. Wandell, "The synthesis and analysis of color images," *IEEE Trans. PAMI* 9 (1): 2–13, 1987.

G. West and M.H. Brill, "Necessary and sufficient conditions for von Kries chromatic adaptation to give color constancy," *J. Math. Biol.* 15: 249–258, 1982.

J.R. Woodwark and K.M. Quinlan, "Reducing the effect of complexity on volume model evaluation," *Comput.-Aided Design J.* 14 (2), 1984.

J.A. Worthey, "Limitations of color constancy," *J. Opt. Soc. Amer. A* 2 (7): 1014–1026, 1985.

J.A. Worthey and M.H. Brill, "Heuristic analysis of von Kries color constancy," *J. Opt. Soc. Amer. A* 3 (10): 1708–1712. 1986.

G. Wyszecki and W.S. Stiles, *Color Science: Concepts, and Methods, Qualitative Data and Formulae.* J. Wiley: New York, 1982.

G. Wyszecki and W.S. Stiles, "High-level trichromatic colour matching and the pigment bleaching hypothesis, *Vision Research* 20: 23–37, 1980.

## Appendix I

PROPOSITION. The set of maps that take a single vertex $p$ of the observed gamut to a point inside the observed canonical gamut, $\mathcal{C}_o$, is a convex polyhedron, $\mathfrak{M}_p$, whose vertexes are the maps that take $p$ to the vertexes of $\mathcal{C}_o$.

*Proof:* If $f(p) \in \mathcal{C}_o$ and $g(p) \in \mathcal{C}_o$, then so is $(1 - \mu)f(p) + \mu g(p)$ for $0 \leq \mu \leq 1$, by the convexity of $\mathcal{C}_o$. Thus, $\mathfrak{M}_p$ is convex. Clearly the maps that take $p$ to some vertex of $\mathcal{C}_o$ are on the boundary of $\mathfrak{M}_p$. Consider the vertexes, $v_i$, of a face of $\mathcal{C}_o$, and maps $F_i(p)$, with $F_i(p) = v_i$. Clearly $F_i$ lies on the boundary of $\mathfrak{M}_p$. Furthermore, if the set $(1 - \mu)v_i + \mu v_j$, $\forall \mu \in [0, 1]$ is an edge of $\mathcal{C}_o$, then $(1 - \mu)F_i + \mu F_j$ lies on the boundary of $\mathfrak{M}_p$. One shows that the boundary of $\mathfrak{M}_p$ consists of plane faces, edges, and vertexes in this way. Thus, $\mathfrak{M}_p$ is a convex polyhedron.

## Appendix II

PROPOSITION. For $\Sigma \subset R^L$, and $F, G \in Gl(L, \mathbf{R})$

$$F(\Sigma) = G(\Sigma) \Leftrightarrow G = FA$$

for some $A \in Gs[\Sigma]$, where $Gs[\Sigma]$ is the linear similarity group of $\Sigma$.

*Proof:* $\Leftarrow$ is clearly true. Now $F, G \in GL(L)$, and so $F$ and $G$ are bijective, by definition. Then

$$G = FA$$

and for some $A$. In fact,

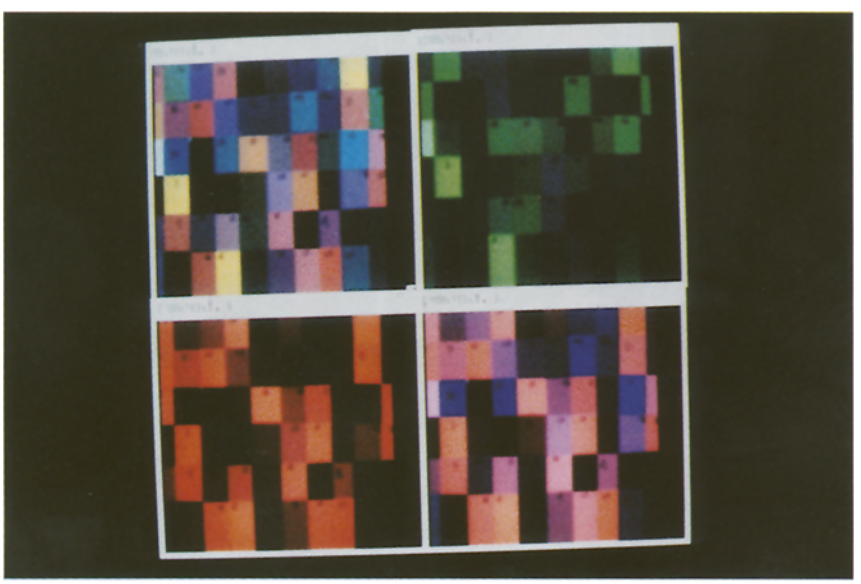
$$A = F^{-1}G$$

but since $F(\Sigma) = G(\Sigma)$, and $F$ and $G$ are bijective, $F^{-1}G(\Sigma) = \Sigma$. Thus $\Rightarrow$ is true.

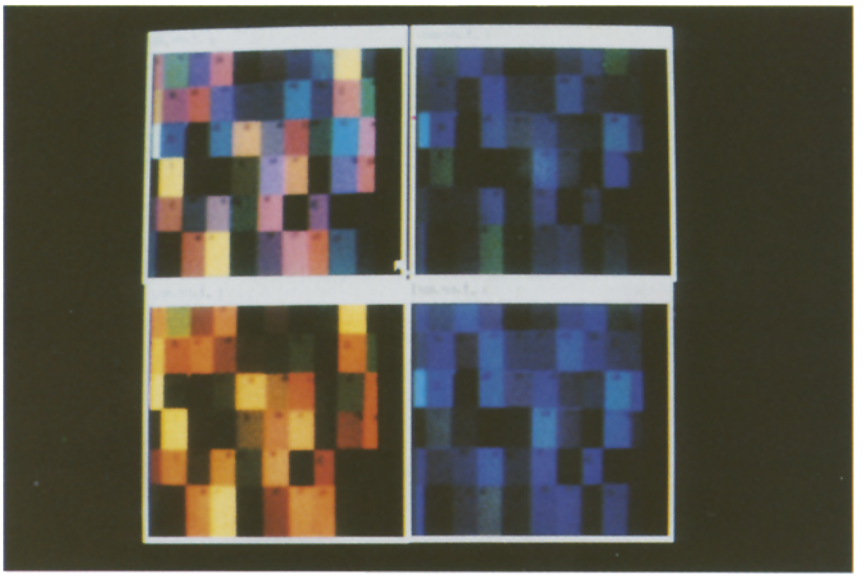
PROPOSITION. For $\Sigma \subset \mathbf{R}^L$, $F \in GL(L)$, $Gs[\Sigma]$ and $Gs[F(\Sigma)]$ are isomorphic.

*Proof:* Consider $f: Gs[\Sigma] \rightarrow Gs[F(\Sigma)]$ corresponding in the natural way to $F : \Sigma \rightarrow F(\Sigma)$. Clearly, for $A \in Gs[\Sigma]$, $f(A) = FAF^{-1} \in Gs[F(\Sigma)]$. We verify easily that the map $f$ is a homomorphism, for $FI_nF^{-1} = I_n$, and $FABF^{-1} = FAF^{-1}FBF^{-1}$. In fact, the map is injective, for $FAF^{-1} = FBF^{-1} \Leftrightarrow A = B$. But we may define a similar injective homomorphism $\bar{f}$: $Gs[F(\Sigma)] \rightarrow Gs[\Sigma]$ by $\bar{f}(B) = F^{-1}BF$. Hence, the groups are isomorphic.

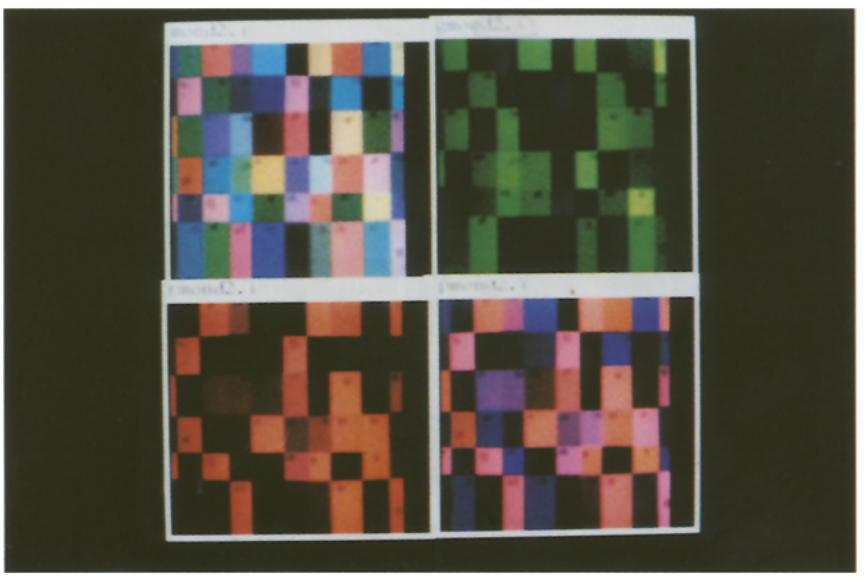
# COLOR FIGURES



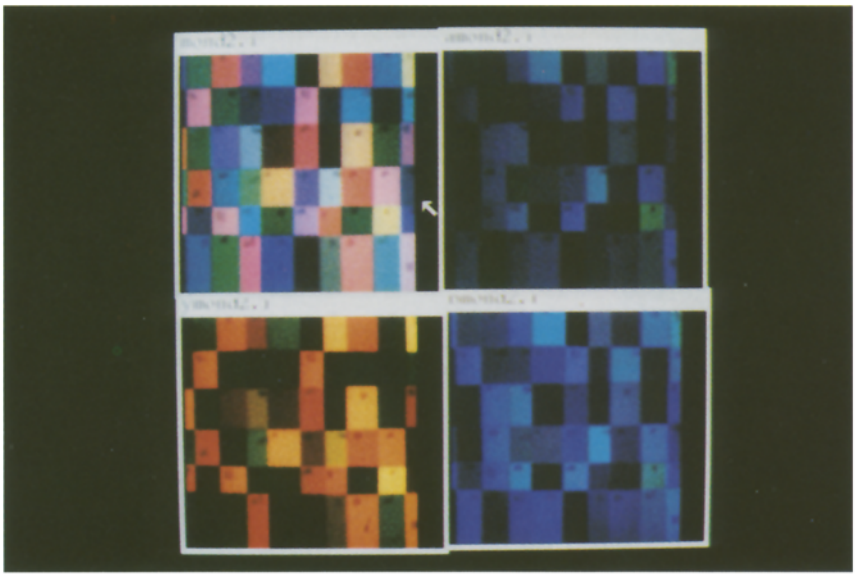
*Color Figure 1.* Images of the first Mondriaan, imaged under (clockwise, from top left) white, green, purple, and red lights. Note the enormous differences. These images show receptor responses. In each case, a human observer was able to name the colors of the Mondriaan patches.



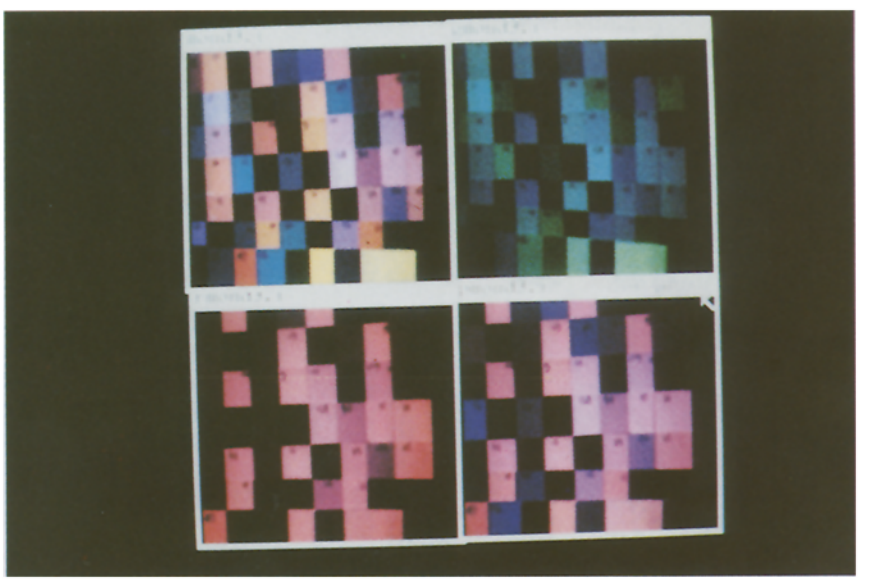
*Color Figure 2.* Images of the first Mondriaan, imaged under (clockwise, from top left) white, blue-green, blue, and yellow lights.



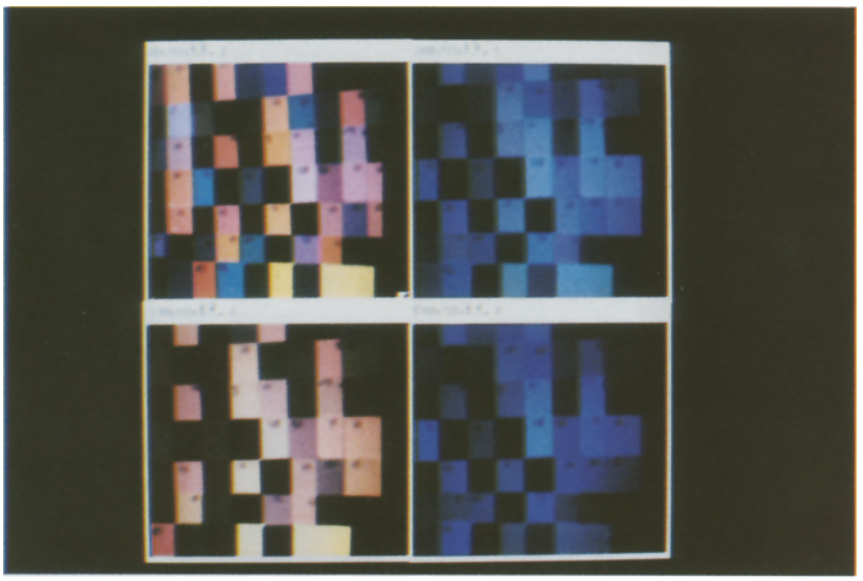
*Color Figure 3.* Images of the second Mondriaan, imaged under (clockwise, from top left) white, green, purple, and red lights.



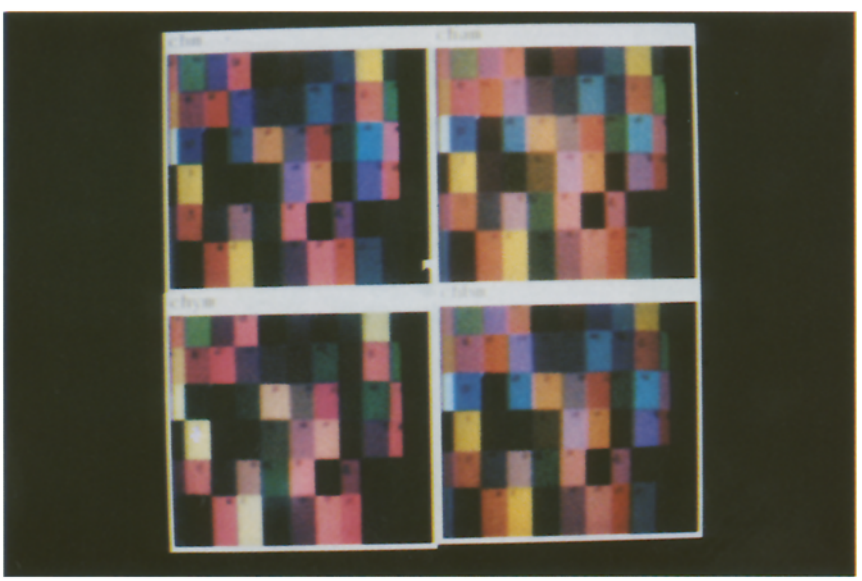
*Color Figure 4.* Images of the second Mondriaan, imaged under (clockwise, from top left) white, blue-green, blue, and yellow lights



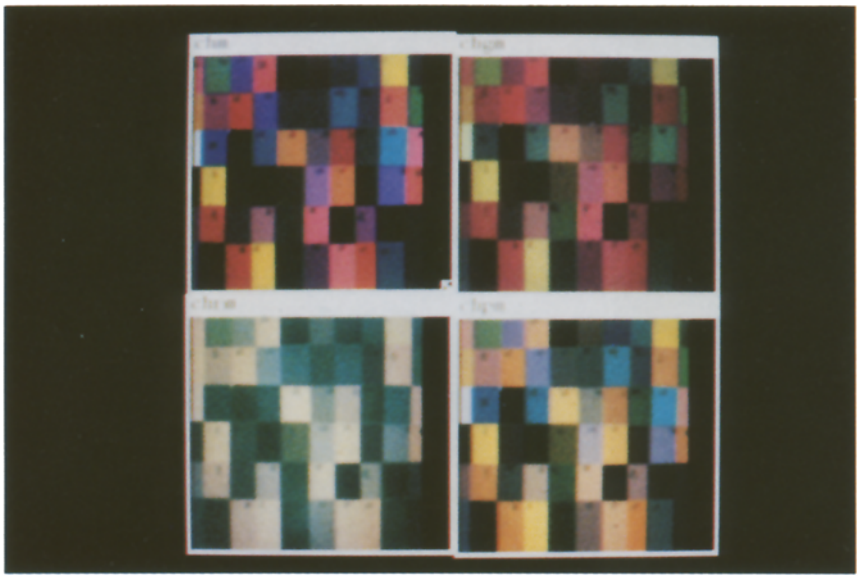
*Color Figure 5.* Images of third Mondriaan, imaged under (clockwise, from top left) white, green, purple, and red lights.



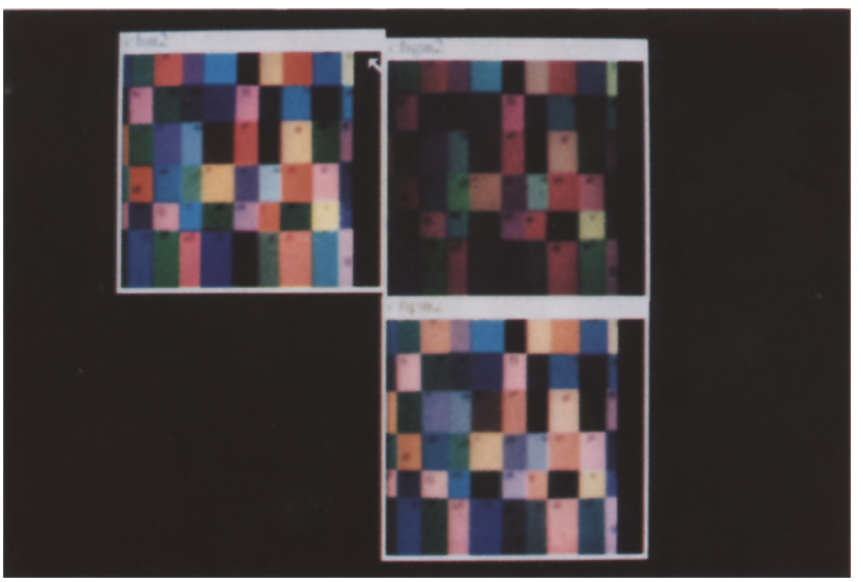
*Color Figure 6.* Images of third Mondriaan, imaged under (clockwise, from top left) white, blue-green, blue, and yellow lights.



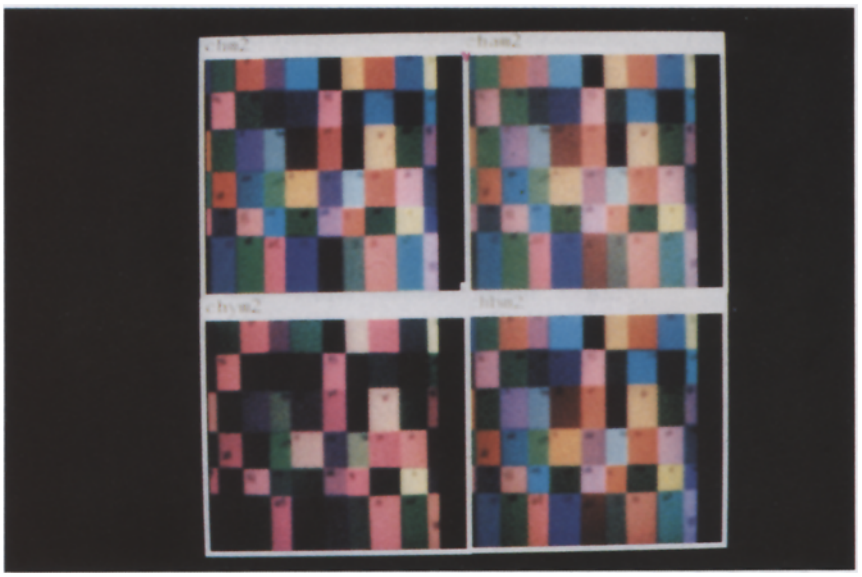
*Color Figure 7.* Outputs of **Crule** for the first Mondriaan derived from images under (clockwise, from top left) white, blue-green, blue, and yellow lights. **Crule** performs well, because these images look similar.



*Color Figure 8.* Outputs of **Crule** for the first Mondriaan derived from images under (clockwise, from top left) white, green, purple, and red lights. Notice that **Crule** performs poorly on the image taken under red light. This effect is explained in the text.

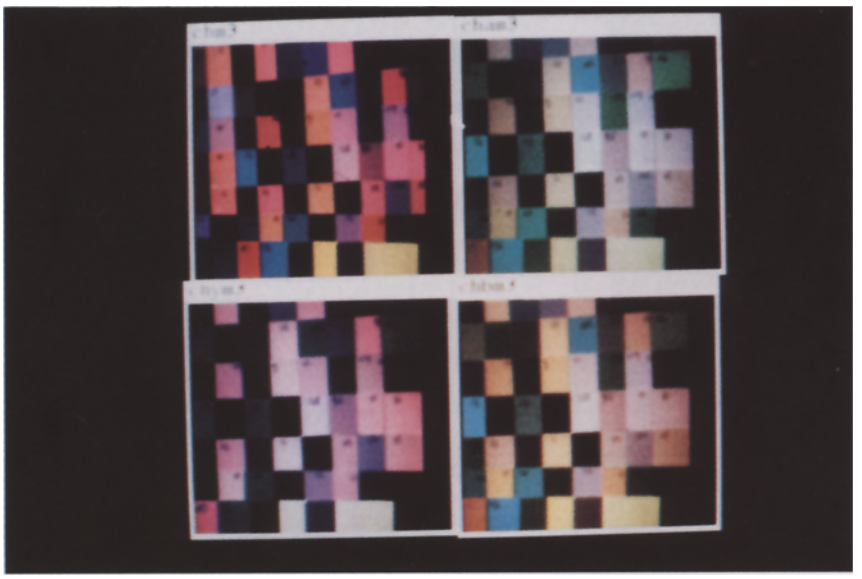


*Color Figure 9.* Outputs of **Crule** for the second Mondriaan derived from images under (clockwise, from top left) white, green, and purple lights. Again, **Crule** performed poorly on the image taken under red light, which was omitted for this reason.

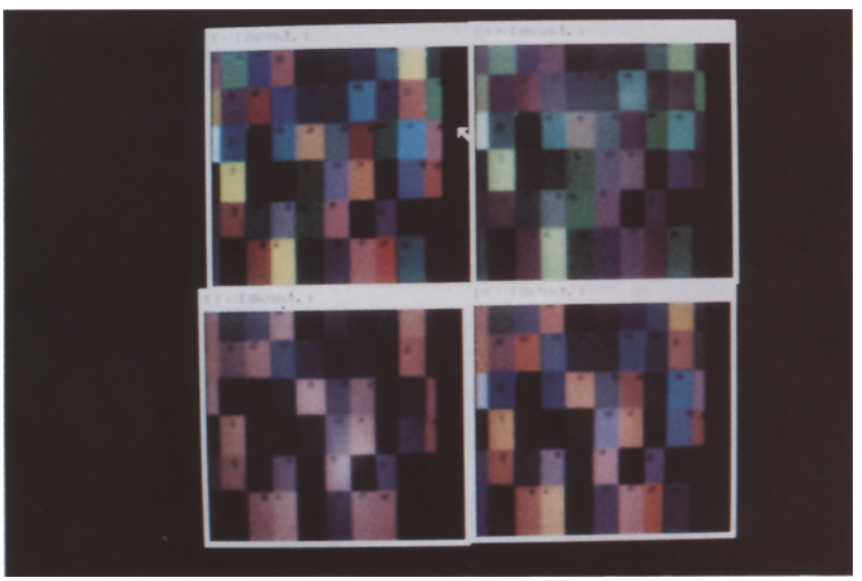


*Color Figure 10.* Outputs of **Crule** for the second Mondriaan derived from images under (clockwise, from top left) white, blue-green, blue, and yellow lights.

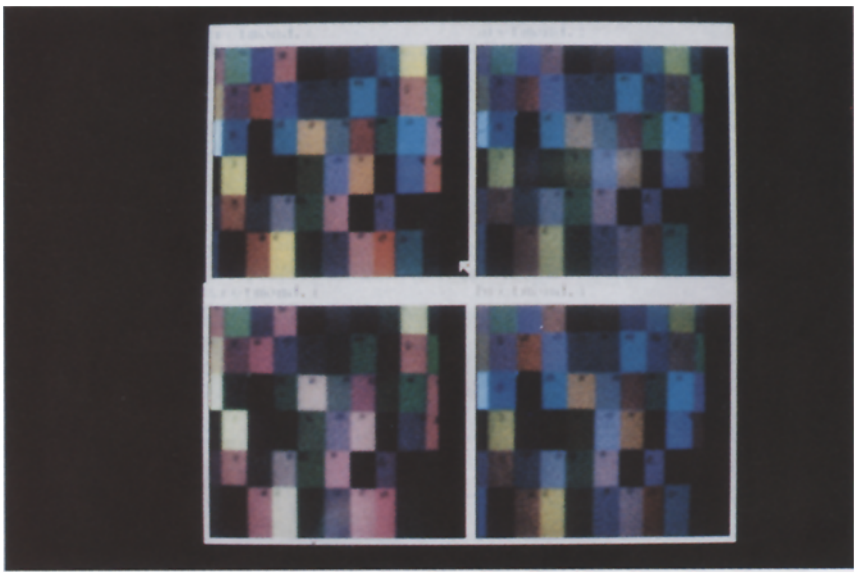
*Color Figure 11.* Outputs of **Crule** for the third Mondriaan derived from images under (clockwise, from top left) white, green, purple, and red lights. **Crule** is not performing very well on these images, because they do not contain a wide range of colors.



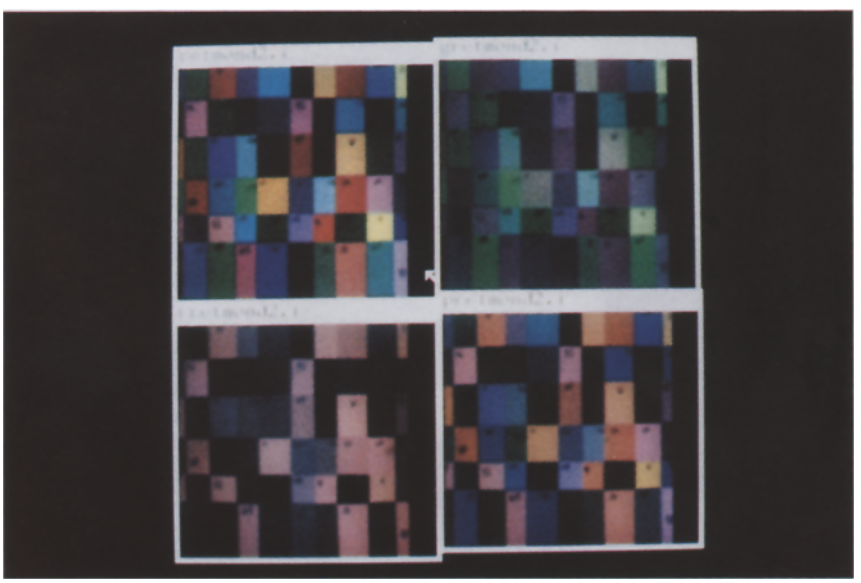
*Color Figure 12.* Outputs of **Crule** for the third Mondriaan derived from images under (clockwise, from top left) white, blue-green, blue, and yellow lights. **Crule** is not performing very well on these images, because they do not contain a wide range of colors.



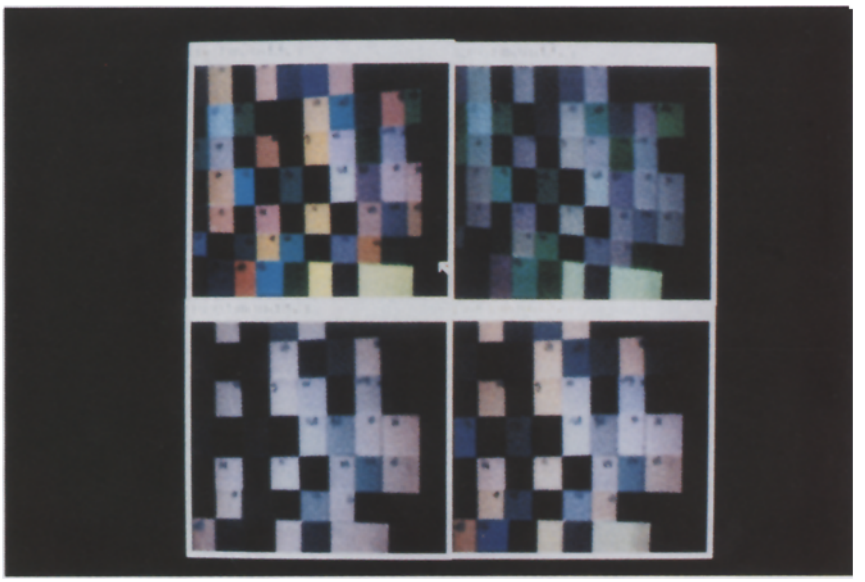
*Color Figure 13.* Outputs of the Brainard and Wandell (1986) model of the Retinex algorithm for the first Mondriaan derived from images under (clockwise, from top left) white, green, purple, and red lights.



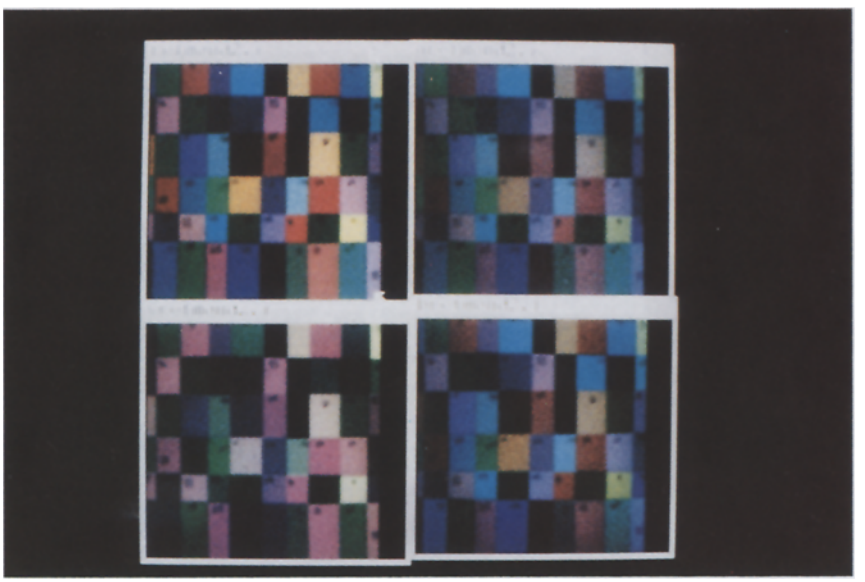
*Color Figure 14.* Outputs of the Brainard and Wandell (1986) model of the Retinex algorithm for the first Mondriaan derived from images under (clockwise, from top left ) white, blue-green, blue, and yellow lights.



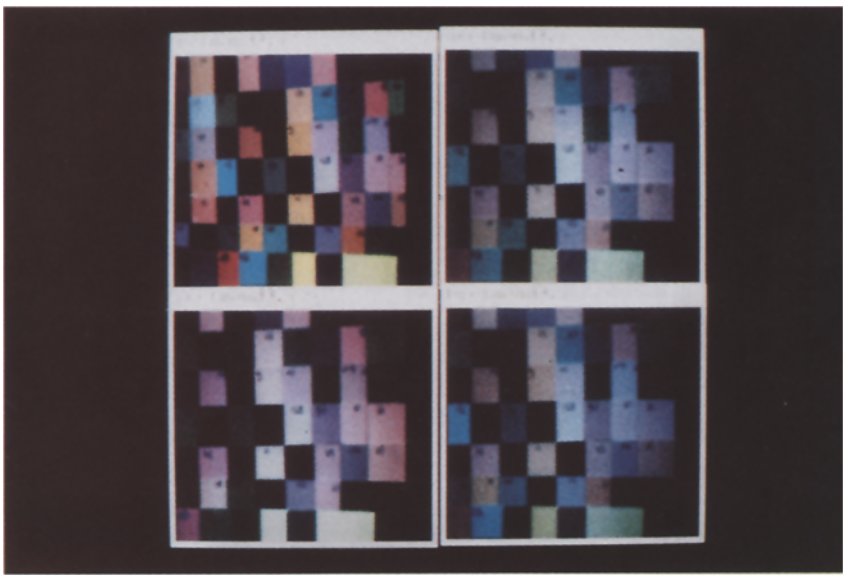
*Color Figure 15.* Outputs of the Brainard and Wandell (1986) model of the Retinex algorithm for the second Mondriaan derived from images under (clockwise, from top left) white, green, purple, and red lights.



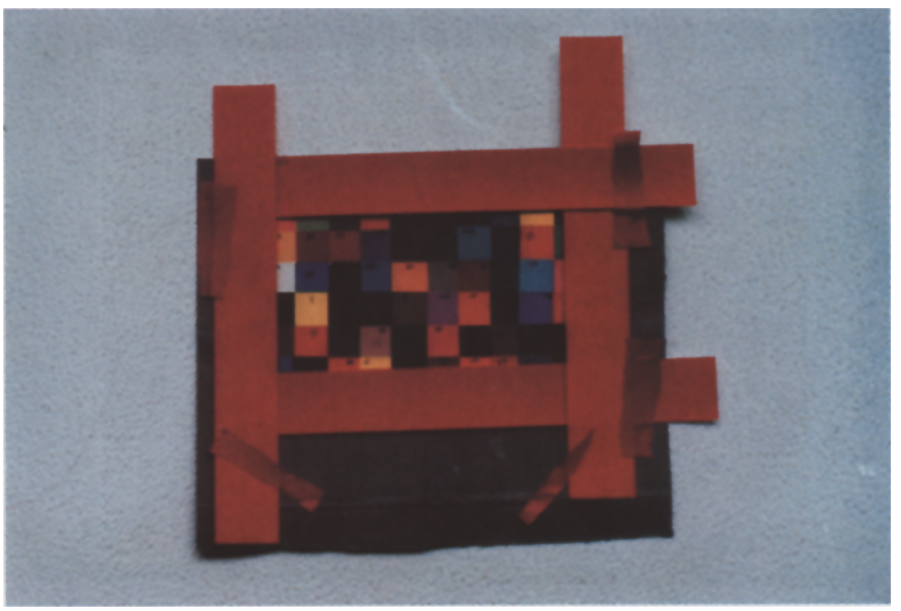
*Color Figure 16.* Outputs of the Brainard and Wandell (1986) model of the Retinex algorithm for the second Mondriaan derived from images under (clockwise, from top left) white, blue-green, blue, and yellow lights.



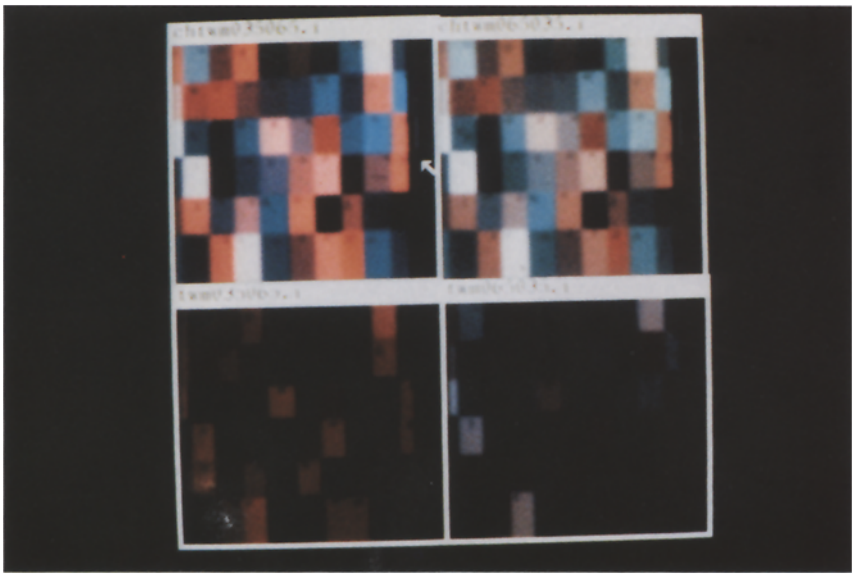
*Color Figure 17.* Outputs of the Brainard and Wandell (1986) model of the Retinex algorithm for the third Mondriaan derived from images under (clockwise, from top left) white, green, purple, and red lights.



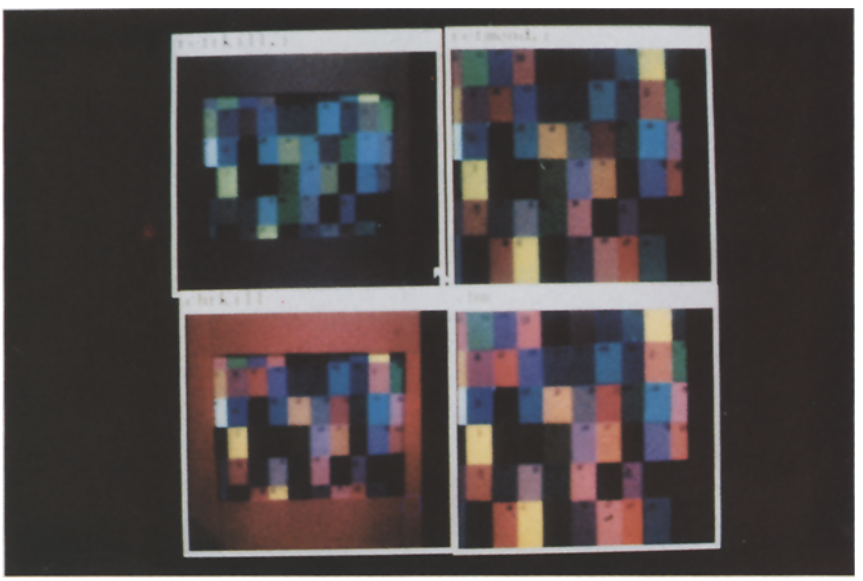
*Color Figure 18.* Outputs of the Brainard and Wandell (1986) model of the Retinex algorithm for the third Mondriaan derived from images under (clockwise, from top left) white, blue-green, blue, and yellow lights.



*Color Figure 19.* Photograph of first Mondriaan with red borders attached.



*Color Figure 20.* Top: The outputs computed by the Retinex algorithm, for an image of the first Mondriaan under white light (right), and for an image of the first Mondriaan with red borders (left), again under white light. The greenish appearance of the left-hand image is caused by the Retinex algorithm confusing a change in average surface color with a colored illuminant. Bottom: Outputs computed by **Crule** for an image of the first Mondriaan under white light (right), and for an image of the first Mondriaan with red borders (left), again under white light. Because **Crule** does not use average surface color, appending borders to the Mondriaan does not affect the color descriptors it computes.



*Color Figure 21.* **Crule** outputs for two separate two color images. The bottom two images are the input images, the top the output images. The image on the left is for a red filter which passed very much more light than that on the right. Details appear in section 7.