MASSACHUSETTS INSTITUTE OF TECHNOLOGY

ARTIFICIAL INTELLIGENCE LABORATORY

WORKING PAPER 207 December 1980

EXACT REPRODUCTION OF COLORED IMAGES

Berthold K. P. Horn

ABSTRACT

The problem of producing a colored image from a colored original is analyzed. Conditions are determined for the production of an image, in which the colors cannot be distinguished from those in the original by a human observer. If the final image is produced by superposition of controlled amounts of colored lights, only a simple linear transform need be applied to the outputs of the image sensors to produce the control inputs required for the image generators. In systems which depend instead on control of the concentration or fractional area covered by colored dyes, a more difficult computation is called for. This calculation may for practical purposes be expressed in table look-up form.

The conditions for exact reproduction of colored images should prove useful in the design and analysis of image processing systems whose final output is intended for human viewing. Judging by the design of many existing systems, these rules are not generally known or adhered to. Modern computational techniques make it practical to tackle this problem now. Adherence to design constraints developed here is of particular importance where colors are to be judged when the original is not directly accessible to the observer as, for example, when it is on another planet.

INTRODUCTION

Systems for the reproduction of colored images have evolved considerably since they were first invented and several hold the potential for accurate reproduction of a major portion of the gamut of possible colors [ ]. Such systems may be conveniently thought of as consisting of the following parts: a set of image sensors exposed to the original image, a set of image generators producing the final image and a computational subsystem mapping the image sensor outputs into suitable inputs for the image generators (see Fig. 1). The image sensors may be photoelectric devices or compounds which undergo chemical changes when exposed to light. The image generators may be controlled light-sources or phosphors, or light-absorbing substances whose concentration or fractional area coverage is controlled. The computational subsystem may be nothing more than a direct coupling between photosensitive substances and other chemicals which can be developed into light-absorbing dyes. It will be shown however that in all but the simplest cases the computations to be performed are more complex than those which such a simple system is able to support. The availability of modern electronic and digital techniques provides us with the tools required to overcome the obstacle presented by the limitations of the straightforward analog or "chemical" computation.

Since such computational techniques did not exist when presently used methods of reproduction were developed, we have come to accept the limited fidelity possible with simpler schemes. It is also often the case that color reproductions are judged more on their appeal to the viewer than on their faithfulness, particularly since the original is not usually available to permit detailed comparisons. Most systems for the reproduction of colored images do obey two fundamental rules nevertheless: the system must have three types of image sensors (with linearly independent spectral response curves) and three types of image generators (again with linearly independent spectral curves). These rules reflect the trichromacy of human color vision, which is illustrated by our ability to match an arbitrary colored light with one made by addition of varying proportions of three test lights [ ].

This observation leads to the assumption that humans possess three types of light-sensitive receptors, presumably the cones in the retina, with linearly independent spectral response curves. These curves are quite similar for a large fraction of the population, with a few exceptions, where one of the three sets appears to be non-functional and a few even rarer cases where one of the sets of receptors has altered spectral response curves [ ]. Experiments further show that these response curves are remarkably stable and that their general shape is unperturbed by adaptation or overexposure. That is, colors may appear different when viewed with the eye adapted differently, but color matching is not disturbed.

PREVIEW

The problem of the reproduction of colored images is analyzed. Conditions are determined for the production of an image in which the colors cannot be distinguished from those in the original by a human observer. By taking point-by-point equality of stimulation of receptors as the criterion of indistinguishability, complicated questions of human color perception are avoided, and it is shown that the spectral response curves of the image sensors must be linear transforms of the spectral response curves of the human visual system.

Having established this design constraint on the image sensors, the computation of control inputs for the image generators is studied next. This computation depends strongly on the method chosen for producing the final image. If, for example, a set of controlled light-sources is superimposed, as in color television, it turns out that a linear transform of the image sensor signals is all that is required. This transform is exhibited as a function of the properties of the human visual system, the system's sensors and the light-sources.

In practice, the range of control of the image generators is limited. In the present case, for example, there is clearly a constraint imposed by the impossibility of negative light intensities. When manipulating absorbing dyes, one is similarly limited to non-negative absorption values. In both cases such constraints lead to limitations on the gamut of colors which can be accurately reproduced. This gamut can be extended by using more than three image generators. The calculations needed to deal with this case are also developed. It is shown that techniques for the solution of linear programming problems are appropriate.

Color reproduction techniques depending on light-absorbers rather than light-sources are analyzed next. These include ordinary photographic processes, where the concentration of dyes in superimposed layers is controlled, and lithographic methods, where the fractional areas covered by dyes are manipulated to achieve the desired effect. It is shown that in general the computations are quite complex, unless unrealistic assumptions are made about the spectral curves of dye absorption. In particular, photographic techniques do not permit the required cross-coupling between layers. That is, each sensitive layer controls only one dye layer in the reproduction. Similarly, color separation and masking techniques for lithographic reproduction cannot cope with the non-linearities due to superposition of non-ideal inks. Curiously, it is usually claimed that masking is required to deal with ink imperfections. The exact calculations proposed here need only be carried through once and the results can then be saved as a three-dimensional look-up table. This table produces the correct control inputs for the image generators so that they give rise to the desired stimulation in the human observer.

When the observer views the reproduction under conditions of adaptation different from those under which he might view the original, it becomes necessary to adjust the system so he will still be able to correctly judge colors. The proper point for this adjustment is identified in the system, based on a simple model of the effect of chromatic adaptation. In traditional systems, adjustments for these effects are introduced at somewhat arbitrary points.

Many of the images we view, such as color television pictures, reproductions in magazines and motion picture film, have been through many reproduction steps. It is therefore important to understand this duplication process. Naturally, systems that accurately reproduce arbitary images will also correctly reproduce reproductions. It is shown, however, that this task is simpler. In particular, it turns out that the sensor spectral response curves for such systems need not be linear transforms of the human spectral response curves.

Many different methods have been used for the reproduction of colored images (see fig. 2). These can be categorized in a number of ways. One can distinguish between those which use controlled intensities of superimposed light-sources with different spectral distribution (as in color television) and those which depend on controlled amounts of absorption by pigments or dyes. The latter can be further divided into a group which requires light to be transmitted through the absorbing layers (as for photographic transparencies) and a group which depends on light reflected from a substrate (as in lithographic reproduction). Along another dimension, one can distinguish methods which depend on addition of lights and others which depend on multiplication of absorption values when dye layers are superimposed or pigments are mixed (also called "subtractive" mixtures). In the latter case, one can further separate methods according to whether control is achieved by means of changes in the concentration or amounts of dye or whether different colors are obtained instead by varying the fraction of the total area covered by a dye of fixed composition. Only three cases are analyzed in any detail in this paper. In order to make it possible to easily generalize to the other techniques, however, one method was chosen from each column and each row in figure 2.

IMAGES, TRANSPARENCIES AND PRINTS

An image can be thought of as a two-dimensional distribution of light intensity. Similarly, a transparency corresponds to a two-dimensional distribution of light transmission, while a print can be modelled as a two-dimensional distribution of light reflectance. These three variables -- intensity, transmission and reflectance -- obviously are also functions of wavelength. The three reproduction systems analyzed in detail in this paper have been carefully chosen so that each of the above cases is represented. However, even though the end product of a reproduction process may be a material entity such as a transparency or print, it is the <u>image</u> on the observer's retina which produces the stimulation of the receptors in his visual system. Consequently the light used to illuminate the transparency or print must be taken into account. It is simpler then to use as the common denominator in all these discussions this final image and to discuss the reproduction of colored <u>images</u>, rather than the other types of end products. In fact it is impossible to make reproductions which will be indistinguishable from the original under all possible illuminating conditions without actually duplicating the exact spectral curves in the original. It is thus important to specify the lighting of the reproduction for which the reproduction is meant to be exact.

A MODEL FOR A COLOR REPRODUCTION SYSTEM

The image reproduction system consists of a set of image sensors viewing the original, a set of image generators producing the final image and a computational sub-system mapping images sensor outputs into image generator inputs (see fig. 1). There are three types of image sensors with different spectral response curves and three image generators with different spectral output curves.

Since we are aiming at point-by-point equality of stimulation, we can concentrate on a particular point in the original and the corresponding point in the reproduction. Let the spectral distribution of light intensity in the original be $s(\lambda)$; that is, the power emitted from an area δA in a spectral band of width $\delta\lambda$ centered at wavelength λ is $s(\lambda)\ \delta A\ \delta\lambda$. Let the spectral distribution of light intensity at the corresponding point in the reproduction be $o(\lambda)$.

It is important to note that it is not necessary for $o(\lambda)$ to equal $s(\lambda)$ for every wavelength λ. All that is needed is that these two spectral distributions be metameric, that is, indistinguishable to the human observer. In other words, they should produce the same stimulation levels in the three types of light sensitive receptors of the human visual system. Now suppose that the spectral response curves of the observer's visual system are $e_1(\lambda)$, $e_2(\lambda)$ and $e_3(\lambda)$. Then the stimulation levels in the three types of sensors will be equal to E_1, E_2 and E_3, defined as follows:

$$E_i = \int_{\lambda_0}^{\lambda_1} s(\lambda)\, e_i(\lambda)\, d\lambda \qquad \text{for } i = 1, 2, 3$$

Here λ_0 and λ_1 are the limits of the visible part of the electromagnetic spectrum. If we call the corresponding stimulation levels when viewing the reproduction E_i', then we must design the system so that $E_i' = E_i$ (for $i = 1, 2, 3$) for all possible input spectral distributions $s(\lambda)$.

VECTORS -- A USEFUL NOTATION

In what follows it will be convenient to think of various spectral distributions as vectors in an infinite dimensional vector space V. To introduce this idea, imagine that we have measured $s(\lambda)$ say, in each wavelength interval of width 10 nm (namometer), between 380 nm and 760 nm. The resulting 38 numbers can be thought of as components of a vector. Different spectral distributions correspond to different vectors. If we increase the resolution of of our measurements, we approach the situation where the vector can be imagined to have infinitely many components [pg. 175, 5; pg 81, 6]. (We may think of the "components" of such a vector as the values of the spectral distribution at particular wavelengths).

We will restrict our attention here to continuous sensor spectral response curves, not only becuase these are the distributions found in practice, but because they simplify the mathematics. In particular, we do not then have to deal with sensor spectral reponse curves that have zero integral over the visible range of wavelengths, yet are non-zero for some wavelengths in this range. As a result, we can avoid repeated use of phrases of the form "equal almost everywhere", when referring to functions which differ only by a "trivial function" [pg. 83, 6]. A number of useful theorems which normally apply only to finite dimensional vector spaces also apply in this case.

INNER PRODUCTS

If we know $e_i(\lambda)$ say, for the same wavelength intervals as those used above to introduce the idea of an infinite dimensional vector space, the integral for the stimulation levels of the three types of receptors in the human visual system can be approximated by a sum of products,

$$E_i \approx \sum_{k=0}^{37} s(385 + k \times 10)\, e_i(385 + k \times 10)$$

This sum has the familiar form of an inner- or dot-product of the corresponding vectors. Once again we may increase the resolution. As we do this, the sum approaches the integral given in the previous section, and we can therefore conveniently think of this integral as the dot-product of two infinite dimensional vectors [pg. 152, 4; pg. 175, 5],

$$E_i = \int_{\lambda_0}^{\lambda_1} s(\lambda)\, e_i(\lambda)\, d\lambda = \underset{\sim}{s} \cdot \underset{\sim}{e}_i$$

Similarly, if the spectral response curves of the image sensors are $r_1(\lambda)$, $r_2(\lambda)$ and $r_3(\lambda)$, we can express their outputs R_1, R_2 and R_3 as

$$R_i = \int_{\lambda_2}^{\lambda_1} s(\lambda)\, r_i(\lambda)\, d\lambda = \underset{\sim}{s} \cdot \underset{\sim}{r}_i$$

This shorthand notation will simplify the determination of the conditions required for exact reproduction of arbitrary images. For the moment we will ignore the fact that spectral distributions will be non-negative for all wavelength and that consequently not all points in the vector space V correspond to realizable spectral distributions or possible sensor response curves.

Most of the mathematical tools we need are available in discussions of "Inner Product Spaces" [4,7], "Euclidian Vector Spaces" [5], "Function Spaces" [6] and "Hilbert Spaces" [8]. Since the basic results needed are not available in the form required here however, they are derived in the appendix.

ORTHOGONAL COMPLEMENTS

Consider all spectral response curves that can be made from linear combinations of the spectral response curves of the image sensors. Clearly this set, when the spectral response curves are viewed as vectors in V, forms a subspace of V. This subspace will be called S_r, and it is spanned by the set of basic vectors $\{r_i\}$.

$$S_r = \{\underset{\sim}{r} | \underset{\sim}{r} = \sum_{i=1}^{3} a_i \underset{\sim}{r}_i ; \varepsilon a_i \ R\}$$

where R is the set of real numbers. The subspace s_r is clearly three-dimensional, since there are three degrees of freedom in choosing the coefficients a_1, a_2 and a_3. Proceeding in a similar fashion, we can define a three-dimensional subspace S_e, spanned by the set of basic vectors $\{e_i\}$. We will later show that these two subspaces must be identical.

A very useful notion in this regard is that of perpendicularity. Two vectors are considered orthogonal if their dot-product is zero. Now consider a vector $\underset{\sim}{v}$ which is orthogonal to all vectors in S_r. We may say that the vector $\underset{\sim}{v}$ is orthogonal to the subspace S_r. This motivates the definition of the orthogonal complement, V_r, say, composed of all the vectors orthogonal to S_r. One writes $V_r = S\frac{1}{r}$. The orthogonal complement is also a subspace of V, it is however infinite dimensionally, quite unlike S_r. It is shown in the appendix (lemma 1) that,

$$V_r = \{\underset{\sim}{v} \mid \underset{\sim}{v} \cdot \underset{\sim}{r}_i = 0, \text{ for each of } i = 1, 2, 3\}$$

This then is the subspace of all spectral distributions which produce zero outputs from each of the image sensors. Note that it is only because we have allowed negative components in spectral distributions that this subspace is non-trivial (that is, contains any but the zero vector). Proceeding in a similar fashion, we can define V_e, the subspace orthogonal to S_e, containing all the spectral distributions which produce no stimulation in the observer's visual system.

CONSTRAINTS ON THE IMAGE SENSORS

Before we consider the image generators and the computational sub-system, we must decide whether or not the image sensors may have arbitrary spectral response curves. It is quite clear that if the image sensors have the same spectral response curves as those in the observer's visual system, then colors which are metameric will produce equal outputs in the image sensors. Similarly, colors which produce the same outputs from the image sensors cannot be distinguished by the observer. This however is more restrictive than needed, since the same result holds if the image sensor's response curves are linear transforms of the spectral response curves of the observer. That is, if

$$r_j(\lambda) = \sum_{i=1}^{3} a_{ij}\, e_i(\lambda) \qquad \text{for } j = 1, 2, 3$$

Or,

$$(\underset{\sim}{r}_i)^T = A\,(\underset{\sim}{e}_i)^T$$

Where A is the matrix (a_{ij}), and the notation $(x_i)^T$ is used to describe a column vector with three components. This then is a sufficient condition that enough information has been captured to be able to produce the final image. But is it necessary?

THEOREM 1: The spectral response curves of the image sensors must be linear transforms of the spectral response curves of the human visual system.

Proof: If the system is to accurately reproduce arbitrary colored images, then spectral distributions which a human can distinguish must produce different outputs from the image sensors. Conversely, spectral distributions which cannot be distinguished from the outputs of the image sensors, must be metameric, that is, indistinguishable as far as the observer is concerned. If we call the two spectral distributions $\underset{\sim}{s}_1$ and $\underset{\sim}{s}_2$, we have,

$$\underset{\sim}{s}_1 \cdot \underset{\sim}{r}_i = \underset{\sim}{s}_2 \cdot \underset{\sim}{r}_i \quad \text{(all i) implies} \quad \underset{\sim}{s}_1 \cdot \underset{\sim}{e}_i = \underset{\sim}{s}_2 \cdot \underset{\sim}{e}_i \quad \text{(all i)}$$

Now let $\underset{\sim}{s} = \underset{\sim}{s}_1 - \underset{\sim}{s}_2$, then,

$$\underset{\sim}{s} \cdot \underset{\sim}{r}_i = 0 \quad \text{(all i) implies} \quad \underset{\sim}{s} \cdot \underset{\sim}{e}_i = 0 \text{ (all i)}$$

That is, a vector perpendicular to $\underset{\sim}{r}_1$, $\underset{\sim}{r}_2$ and $\underset{\sim}{r}_3$, must also be perpendicular to $\underset{\sim}{e}_1$, $\underset{\sim}{e}_2$ and $\underset{\sim}{e}_3$. Stated another way, any vector $\underset{\sim}{v} \in \underset{\sim}{V}_r$ must also be in V_e. or $V_r \subset V_e$. Now, according to lemma 2 in the appendix, this implies that $V_e^{\perp} \subset V_r^{\perp}$. Further, if sensor spectral response curves are continuous, lemma 6 in the appendix shows that $V_r^{\perp} = S_r$ and $V_e^{\perp} = S_e$. Therefore $S_e \subset S_r$.

At this point we note that the two subspaces S_e and S_r are of equal dimension, and by lemma 7 in the appendix this implies that actually $S_e = S_r$. We have two bases for this vector-space, $\{\underset{\sim}{e}_i\}$ and $\{\underset{\sim}{r}_i\}$. There must then exist a linear transform between these two sets of vectors [pg. 119, 7]. We may represent this linear transformation by means of a matrix, A. Then,

$$(\underset{\sim}{r}_i)^T = A(\underset{\sim}{e}_i)^T$$

Note: It should be apparent that the decision to permit spectral distributions with negative components simplified the derivation, because it permitted the representation of <u>differences</u> of spectral distributions.

<u>COROLLARY 1:</u> Metameric spectral distributions produce identical outputs in the camera.

Proof: Since $S_e = S_r$, we have $S_r \subset S_e$, and therefore $V_e \subset V_r$. That is,

$$\underset{\sim}{s}\cdot\underset{\sim}{e}_i = 0 \quad (\text{all } i) \quad \text{implies} \quad \underset{\sim}{s}\cdot\underset{\sim}{r}_i = 0 \quad (\text{all } i)$$

Note: The spectral response curves of the human visual system are not themselves known with great accuracy, although a large variety of experiments hint at their general shape [ ]. Extensive color matching experiments have however led to agreement on what are called standard observer curves [ ]. These are constructed in such a way that spectral distributions which match in terms of the standard observer curves will be metameric. Since the C.I.E. standard observer curves represent the average of many experiments with many different subjects, while the human cone response curves are not yet known with the same precision, the following result will be useful:

COROLLARY 2: The C.I.E. standard observer curves are linear transforms of the spectral response curves of the human visual system.

Proof: Immediate, if one replaces the image sensor spectral response curves in the previous theorem with the standard observer curves.

As a result we may restate the constraint: The spectral response curves of the image sensors must be linear transforms of the standard observer curves.

WHAT TO DO IF THE SENSOR RESPONSE CURVES ARE NOT QUITE RIGHT

In practice, it is not possible to design devices with arbitrary spectral response curves. In order to select sensors with known spectral response curves, one would like to know what the "nearest" linear transform of the response curves of the human visual system is and how "near" to it the given response curve is. If one gives a least-squares interpretation to the term "near", the answer is quite simple. The spectral response curve which is a linear transform of the spectral response curves of the human visual system and which is closed to the given response curve, $\underset{\sim}{v}$ say, is the perpendicular projection, $\underset{\sim}{s}$ say, of $\underset{\sim}{v}$ onto S_e. The error is measured conveniently by the perpendicular distance between $\underset{\sim}{v}$ and S_e. As the problem is stated it amounts to minimization of $(\underset{\sim}{v} - \underset{\sim}{s})\cdot(\underset{\sim}{v} - \underset{\sim}{s})$, when $\underset{\sim}{s} \in S_e$. That is, $\underset{\sim}{s}$ can be written as a linear combination of the basis vectors of S_e,

$$\underset{\sim}{s} = \sum_{i=1}^{3} a_i \underset{\sim}{e}_i$$

Using the methods of lemma 3 in the appendix, one finds that $\underset{\sim}{s}$ is the perpendicular projection of $\underset{\sim}{v}$ on S_e and that,

$$(a_i)^T = Q^{-1} (\underset{\sim}{v}\cdot\underset{\sim}{e}_i)^T$$

where the matrix Q has elements $q_{ij} = \underset{\sim}{e}_i\cdot\underset{\sim}{e}_j$. Next, by lemma 4, the vector $(\underset{\sim}{v} - \underset{\sim}{s})$ is perpendicular to S_e and so $(\underset{\sim}{v} - \underset{\sim}{s})\cdot(\underset{\sim}{v} - \underset{\sim}{s}) = \underset{\sim}{v}\cdot(\underset{\sim}{v} - \underset{\sim}{s}) = \underset{\sim}{v}\cdot\underset{\sim}{v} - \underset{\sim}{v}\cdot\underset{\sim}{s}$. When comparing different sensors one should normalize this "error" by the total sensitivity of the sensor. This produces a "quality factor",

$$\underset{\sim}{v}\cdot\underset{\sim}{s}/\underset{\sim}{v}\cdot\underset{\sim}{v}$$

which will be one for a perfect sensor.

A SYSTEM WHICH USES SUPERIMPOSED CONTROLLED LIGHT-SOURCES

Now we turn our attention to the image generators and the computation required between the outputs of the image sensors and the inputs to the image generators. Here, we first consider the simplest system, one which superimposes three colored light-sources whose intensity can be individually controlled. Color television represents the most widely known instance. The light-sources in this case are the phosphor dots on the screen whose intensity is controlled by the current in the incident electron beam. The light-sources, while not actually in the same place, appear essentially superimposed at normal viewing distance.

Let $p_1(\lambda)$, $p_2(\lambda)$ and $p_3(\lambda)$ be the spectral distributions of the light emitted by each of the three light-sources. Further, let P_1, P_2, P_3 represent their absolute levels (these are the control inputs to the image generation system). Then, the total spectral distribution of light coming from a particular point can be found by simple addition,

$$o(\lambda) = P_1 p_1(\lambda) + P_2 p_2(\lambda) + P_3 p_3(\lambda)$$

Or,

$$\underset{\sim}{o} = \sum_{j=i}^{3} P_j \underset{\sim}{p}_j$$

From this we can calculate the stimulation of the receptors in the observer's visual system when viewing the reproduction as follows:

$$E_i' = \underset{\sim}{o}\cdot\underset{\sim}{e}_i = \sum_{j=i}^{3} P_j\,(\underset{\sim}{e}_i\cdot\underset{\sim}{p}_j)$$

This suggests that one defines a matrix $C = (c_{ij})$ say, where $c_{ij} = \underset{\sim}{e}_i\cdot\underset{\sim}{p}_j$. Then,

$$(E_i')^T = C(P_i)^T$$

Note: Since the spectral distributions $p_j(\lambda)$ and the spectral sensitivities $e_i(\lambda)$ are non-negative, all terms in the matrix, c_{ij} are non-negative. This has implications for the inverse of the matrix C. Some elements of the inverse matrix must be negative, for example.

THEOREM 2: The mapping of image sensor outputs to image generator inputs can be achieved by means of a linear transform if A and C are non-singular. This linear transform can be represented by a 3 x 3 matrix $B = (AC)^{-1} = C^{-1}A^{-1}$.

Proof: If the final image is to be indistinguishable from the original, $E_1' = E_1$, $E_2' = E_2$ and $E_3' = E_3$. Now if C is non-singular,

$$(P_i)^T = C^{-1}\,(E_i)^T$$

Finally, one has to find the stimulation levels $(E_i)^T$ from the image sensor outputs $(R_i)^T$. If A is non-singular,

$$(E_i)^T = A^{-1}\,(R_i)^T$$

So,

$$(P_i)^T = C^{-1} A^{-1} (R_i)^T = (AC)^{-1} (R_i)^T$$

Thus the calculation can be performed by a simple linear transform, which can be represented by a 3 x 3 matrix (see fig. 3).

COROLLARY 3: The set of vectors $\{\underset{\sim}{r}_i\}$ must be linearly independent, as must the set $\{p_i\}$.

Proof: It follows from the trichromacy of human color vision that the set of vectors $\{\underset{\sim}{e}_i\}$ must be linearly independent. Now $(\underset{\sim}{r}_i)^T = A(\underset{\sim}{e}_i)^T$, so if the set $\{\underset{\sim}{r}_i\}$ were linearly dependent, the rows of the matrix A would be. The condition that A is non-singular thus implies that the set of <u>image sensor</u> spectral response curves $\{\underset{\sim}{r}_i\}$ be linearly independent.

Next, note that one can write the matrix C as the product of a column vector $(\underset{\sim}{e}_i)^T$ and a row vector $(\underset{\sim}{p}_i)$. If the set $\{p_i\}$ were linearly dependent, so would the columns of C. The condition that C is non-singular thus implies that the set of <u>image generator</u> spectral curves $\{\underset{\sim}{p}_i\}$ be linearly independent.

Note: The linear independence of each of these sets of spectral curves is necessary, but <u>not</u> sufficient. The requirement that the matrices A and C be non-singular is more stringent.

<u>COROLLARY 4:</u> The light-source spectral distributions, $p_i(\lambda)$, need not be linear transforms of the spectral response curves of the human visual system.

Proof: This is clear from the proof of the previous theorem, since the matrix B can be found for <u>arbitrary</u> non-singular matrices A and C.

<u>COROLLARY 5:</u> In determining the transform B, we can use matrices A' and C' based on the standard observer curves, instead of the matrices A and C, based on the actual spectral response curves of the human visual system.

Proof: We have already shown that the C.I.E. standard observer curves represent a linear transform of the spectral response curves of the human visual system. Let us represent this transformation by the non-singular matrix $M = (m_{ij})$ say. Then if we let $e_i'(\lambda)$ be the standard observer curves,

$$(\underset{\sim}{e}_i')^T = M(\underset{\sim}{e}_i)^T$$

Consequently,

$$(\underset{\sim}{r}_i)^T = A(\underset{\sim}{e}_i)^T = A\, M^{-1}(\underset{\sim}{e}_i')^T. \quad \text{So } A' = A\, M^{-1}.$$

Next, note that $C = (\underset{\sim}{e}_i)^T(\underset{\sim}{p}_i)$ and that similarly $C' = (\underset{\sim}{e}_i')^T(\underset{\sim}{p}_i)$. Since $(\underset{\sim}{e}_i')^T = M(\underset{\sim}{e}_i)^T$, $C' = MC$. As a result,

$$(A'C')^{-1} = (A\ M^{-1}\ M\ C)^{-1} = (AC)^{-1} = B$$

This is very convenient, since the standard observer curves have been determined with fair accuracy, while there is continuing debate about the exact form of the spectral response curves of the human visual system. That is, the linear transform between the two has not been pinned down as accurately as one would wish.

<u>COROLLARY 6:</u> The image sensor outputs may be connected directly to the image generator inputs if and only if $AC = I$, the 3 x 3 identity matrix.

Proof: The direct connection implies that $B = I$. The result follows since $B = (AC)^{-1}$.

Notice that this condition is very restrictive, and the exact condition is unlikely to be met in practice if one keeps in mind the limitations imposed on possible image sensor and image generator spectral curves. Nevertheless, some present-day system for the reproduction of colored images (as, for example, photography) use such direction connection and do not permit correction for possible cross-coupling terms! In lithography cross-terms can be taken care of by "masking" in the color separation step. It will be shown later however that other inaccuracies occur in this case.

CHROMATIC ADAPTATION

The human visual system adjusts to varying lighting conditions. While there is as yet no general agreement on the exact nature of the processes involved, or where in the visual system it takes place, adaptation is often modelled as a change in gain of the channels [ ]. So, for example, it is likely that an observer viewing objects illuminated by incandescent light is compensating for the strong illumination in the long wavelength end of the spectrum by means of a reduced gain in the receptor channel most sensitive to long wavelengths and by an increased gain in the receptor channel most sensitive to the short wavelengths.

Systems for the reproduction of colored images may take this effect into account by introducing corresponding gain changes in order to deal with the fact that the viewer is likely to be adapted differently, when viewing the reproduction, than (s)he would be when viewing the original. To some extent this is already done in existing systems in order to deal with the limited range of values available in the image generators. So, for example, in color photography, one would use a film that is "balanced" for the incandescent illumination -- that is, a film that has reduced sensitivity in the long wavelength sensitive layer and increased sensitivity in the short wavelength sensitive layer. The sensitivities in such a film are adjusted so that a white surface illuminated by a tungsten lamp of the specified color temperature will be reproduced as a white or neutral color in the final image. In this way the range of intensities in the original scene can be fitted into the limited dynamic range of the film. Each of the three dye layers in the film (to be described in more detail later) is called upon to produce a similar range of absorbing densities.

Most systems for the reproduction of colored images introduce this gain change into the channels directly connecting image sensors to image generators. This however does not usually produce the correct transformation.

THEOREM 3: The gain changes required to compensate for the observer adaptation level should be introduced between the linear transform A^{-1} and the linear transform C^{-1}.

Proof: If adaptation can be modelled as gain changing in the receptor channels, it can be compensated for by applying the inverse gain changes in the reproduction system. To do this, we must first calculate the receptor stimulation levels from the image sensor outputs.

$$(E_i)^T = A^{-1} (R_i)^T$$

At this point we multiply by the gain-factors g_1, g_2 and g_3. Finally we calculate the image generator control inputs as before,

$$(P_i)^T = C^{-1} G (E_i)^T$$

where G is the diagonal matrix with elements g_1, g_2 and g_3. The two linear transform and the gain-factors are shown graphically in figure 4.

<u>COROLLARY 7:</u> Applying the gain factors to image sensor outputs or image generator imputs will <u>not</u> in general result in correct compensation for adaptation.

Proof: According to the previous theorem, the overall transfer function of the system from image sensor outputs to image generator inputs should be $C^{-1}\, G\, A^{-1}$. If we try to achieve the same effect by modifying the image sensor outputs first, we obtain instead $C^{-1}\, A^{-1}\, G'$, where G' is a new diagonal matrix. If the transfer functions are supposed to be equal, we find that $G\, A^{-1} = A^{-1}\, G'$ or $AG = G'A$.

One can see the impossibility of this in the general case, since the matrix on the left is obtained from A by scaling its <u>columns</u>, while the matrix on the right is obtained by scaling its <u>rows</u>. Put another way, $G' = A\, G\, A^{-1}$, which is <u>not</u> diagonal unless A is diagonal.

The same sort of argument shows that applying gain factors to the image generator inputs will not do, since one then obtains $G''\, C^{-1}\, A^{-1}$ and so require that $C^{-1}G = G''\, C^{-1}$. Therefore one finds $G'' = C\, G\, C^{-1}$ and so has the same difficulties unless C is diagonal (which is not the case because of the overlap of the spectral response curves of the human visual system).

Note: So far we have been able to use the C.I.E. standard observer curves in our derivations. Here however we actually have to get at the underlying receptor response curves, since the gain factors are to be interposed between A^{-1} and C^{-1}. This is quite reasonable, since in fact chromatic adaptation experiments represent one technique for estimating the actual receptor response curves.

CONSTRAINTS ON IMAGE GENERATOR CONTROL INPUTS

For certain image sensor outputs, the calculation presented so far may result in negative control inputs to the image generators. This is an indication that the correct stimulation levels of the human visual system cannot be achieved by adding non-negative amounts of the three light-sources. Since negative intensities cannot be realized, we conclude that the image generators can produce only a limited gamut of observer stimulation levels and that consequently some spectral distributions cannot be reproduced correctly.

<u>THEOREM 4:</u> The set of observer stimulation levels $(E_i)^T$ that can be produced using non-negative light-source levels forms a <u>convex subset</u> of the space of all possible stimulation levels.

Proof: By adding light-source intensities, arbitrary positive linear combinations of stimulation levels can be produced. That is, if

$$(E_i)_1^T = C\ (P_i)_1^T \quad \text{and} \quad (E_i)_2^T = C\ (P_i)_2^T$$

Then

$$a\ (E_i)_1^T + (1 - a)(E_i)_2^T = C\ [a(P_i)_1^T + (1 - a)(P_i)_2^T]$$

The possible stimulation levels thus form a convex set. Obviously, it would be to our advantage to make this sub-set as large as possible. We are limited in our attempt to do this by the next result.

CORROLARY 8: The subset of stimulation levels possible with arbitrary non-negative spectral distribution is itself convex, and bounded by the stimulation levels produced by monochromatic light-sources.

Proof: This follows from the fact that arbitrary non-negative spectral distributions can be thought of as sums of scaled monochromatic spectral distributions. That is,

$$s(\lambda) = \int_{\lambda_0}^{\lambda_1} s(\lambda') \, \delta(\lambda' - \lambda) \, \delta\lambda'$$

The set of stimulation levels that can be produced using three fixed light-sources is clearly a subset of this set. To make it as large a subset as possible, one ought to use monochromatic light-sources if possible, since these produce stimuli lying on the boundary of the convex set. There is however no set of monochromatic light-sources which will make the subset of stimulation values covered by the image generation system equal to the set of all possible stimulation levels (this is a result of the overlap of human spectral response curves).

A very large portion can be covered however by choosing a source from the short end of the spectrum (between 400 nm and 460 nm), one near the middle (between 500 nm and 540 nm) and one near the long wave-length end (betwen 620 nm and 700 nm). Unfortunately, there is a further constraint: at the extremes of the visible region of the spectrum, the eye is relatively insensitive, and large light-source intensities are needed to produce given stimulation levels. For this reason, the phosphors used in color television represent a compromise between a desire to cover as large a gamut of stimulation levels as possible and the need to produce adequate screen brightness [ ].

USING MORE THAN THREE LIGHT-SOURCES

In order to span a larger range of possible stimulation levels, one may chose to use more than three light-sources in the projector. Clearly non-negative intensity levels still create a convex subset of stimulation levels, but this subset can be made larger than it would be with only three light-sources. If we have n light-sources, the vector $(P_i)^T$ of outputs to the projector will have n components and we can write, much as before

$$(E_i')^T = C\ (P_i)^T$$

where C however is no longer square. That is, we have n > 3 unknowns and only three equations. The solution $(P_i)^T$ is clearly non-unique and many light-source amplitude combinations will produce the same stimulation levels.

Any solution can be expressed in terms of a generalized inverse X of the matrix C [pg. 2, B-I]. That is, if C X C = C, then $C(P_i)^T = (E_i')^T$ has a solution only if $C\ X\ (E_i')^T = (E_i')^T$, in which case the general solution is,

$$(P_i)^T = X\ (E_i')^T + (I - X\ C)(Y_i)^T$$

where $(Y_i)^T$ is an arbitrary vector.

To pick a particular one of this set of possible solutions, one may look for the one with the minimum norm, where the norm may be defined as the sum of squares of the image generator inputs or

$(P_i)(P_i)^T$. This solution can be found using the pseudo-inverse [pg.113,] of the matrix C (see lemma 8 in the appendix).

$$(P_i)^T = (C^TC)^{-1} C^T(E_i')^T$$

Or,

$$(P_i)^T = (C^TC)^{-1} C^T A^{-1} (R_i)^T$$

While this solution fits in nicely with our system so far if we simply let $B = (C^T C)^{-1} C^T A^{-1}$, it does not guarantee non-negative outputs to the projector for all points in the convex subset available to us.

Introducing the non-negativity constraints on the image generator control inputs leads naturally to a linear programming problem which will be discussed next.

THEOREM 5: The problem of the determination of suitable image generator control inputs when there are more than three light-sources can be posed as a problem in linear programming. For given stimulation levels of the receptors of the observer, only three of the light-sources need be used at a time.

Proof: Three linear constraints must be satisfied in order for the stimulation of the receptors in the observer's visual system to be correct,

$(E_i)^T = C(P_i)^T$

Further, there are n inequalities of the form $P_i \geq 0$. In order to pick one of the many possible solutions, one may introduce a cost function,

$$\sum_{i=1}^{n} k_i P_i$$

A convenient example would be a cost function equal to the total energy used by the light-sources (i.e., $k_i = 1$ for all i). The solution which minimizes the cost function can be found by standard linear programming techniques [ ].

A <u>feasible</u> solution is any solution which satisfies the linear constraints and the non-negativity constraints. A <u>basic</u> solution is a solution which contains only m non-zero variables, where m is the number of structural constraints (three in this case). An <u>optimal solution</u> is a feasible solution which minimizes the cost-function [pg. 94,]. The fundamental theorem of linear programming states that if there is an optimal solution, then there is a <u>basic</u> optimal solution. This implies that an optimal solution can be found in which only m variables are non-zero. In the situation here this simply means that a given set of stimulation levels of the observer's visual system can be achieved using no more than three of the light-sources at a time.

FINDING AN INITIAL BASIC FEASIBLE SOLUTION

Before one can apply the well-known SIMPLEX method for solving this linear programming problem, one must find some basic feasible solution. Usually, for problems with the constraints given in the form of inequalities, this is straightforward since the "slack variables" can be used. In this situation, however, since the three constraints are equalities, it is necessary to introduce so-called "artificial variables", A_1, A_2 and A_3 [pg 132,]. To obtain the initial basic feasible solution then, one first solves a new linear programming problem of the form,

$$\begin{aligned}
c_{11}\,P_1 + c_{12}\,P_2 + \dots\, c_{1n}P_n + A_1 \qquad\qquad\quad &= E_1 \\
c_{21}\,P_1 + c_{22}\,P_2 + \dots\, c_{2n}P_n \qquad + A_2 \qquad\quad &= E_2 \\
c_{31}\,P_1 + c_{32}\,P_2 + \dots\, c_{3n}P_n \qquad\qquad + A_3 &= E_3
\end{aligned}$$

with cost-function

$$\sum_{i=1}^{3} A_i$$

This problem obviously has basic feasible solution $A_1 = E_1$, $A_2 = E_2$ and $A_3 = E_3$. If the original problem has a feasible solution, minimization of the cost-function will lead to a solution with $A_1 = A_2 = A_3 = 0$ and consequently result in a basic feasible solution of the original problem. One may elect to simply stop at that point, since a set of light-source intensities has been found which will produce the desired stimulation of the receptors in the observer's visual system. Alternatively, one can continue to minimize the new cost function. The overall computation is not very lengthy, since the simplex method takes between m and (3m/2) pivoting steps typically [ ], and here m = 3.

Note further, that in practice image generator inputs also have upper limits, that is, there is some maximum intensity that each light-source can produce. These constraints too can be incorporated in the linear programming formulation quite easily, with come increased computation. It is clear however that the computation of image generator inputs from image sensor outputs is beginning to be a bit more complicated now and can no longer be carried out by direct connections, simply gain factors or even linear generators.

REPRODUCING REPRODUCTIONS

The system described so far will clearly correctly reproduce reproductions, that is, images produced by a similar system. It is possible to do this with a simpler system however, since the input sensors do not now have to deal with arbitrary spectral distributions. In fact, the space of possible input spectral distributions is finite dimensional. It will be shown that a system for duplication need not have spectral response curves which are linear transforms of the spectral response curves of the human visual system.

This is of great practical significance, since many duplication steps typically lie between original and the image finally presented to the viewer. The final result would be poor indeed, if at each stage the image was further degraded by our inability to build image sensors which have response curves that are exactly equal to some linear transform of those of the human visual system. At the same time, this is the root of considerable confusion, since such systems can be designed around any convenient sensor response curves and image generator spectral curves, while the system viewing the original image must be quite special as has been shown.

THEOREM 6: For the reproduction of reproductions, the spectral response curves of the image sensors need not be linear transforms of the spectral response curves of the human visual system.

Proof: Consider an image made by superimposing various amounts of light from three light-sources. Let the spectral outputs of the light-sources be $f_1(\lambda)$, $f_2(\lambda)$ and $f_3(\lambda)$, and their intensity F_1, F_2 and F_3.

Then the spectral distribution of light intensity of the input to our system will be,

$$s(\lambda) = \sum_{i=1}^{3} F_i f_i(\lambda)$$

Consequently, the image sensor outputs will be,

$$R_j = \underset{\sim}{s} \cdot \underset{\sim}{r}_j = \sum_{i=1}^{3} F_i \underset{\sim}{f}_i \cdot \underset{\sim}{r}_j \quad \text{or} \quad (R_i)^T = H(F_i)^T$$

where the matrix H has elements $h_{ij} = \underset{\sim}{r}_i \cdot \underset{\sim}{f}_j$. Similarly

$$E_j = \underset{\sim}{s} \cdot \underset{\sim}{e}_j = \sum_{i=1}^{3} F_i \underset{\sim}{f}_i \cdot \underset{\sim}{e}_j \quad \text{or} \quad (E_i)^T = G(F_i)^T$$

where the matrix G has elements $g_{ij} = \underset{\sim}{e}_i \cdot \underset{\sim}{f}_j$. Clearly then

$$(R_i)^T = H\,G^{-1}\,(E_i)^T \quad \text{or} \quad (E_i)^T = G\,H^{-1}(R_i)^T$$

That is, we can determine $(E_i)^T$ from $(R_i)^T$ even when the $\underset{\sim}{r}_i$ are not linear transforms of the $\underset{\sim}{e}_i$. So the previous analysis applies if one lets $A = H\,G^{-1}$.

COROLLARY 9: If the image presented as input to the reproduction system was originally made on the system used now to reproduce it, then the linear transform takes on a particularly simple form. That is, $B = H^{-1}$.

Proof: In this simple case, $\underset{\sim}{f}_i = \underset{\sim}{p}_i$, and so $G = C$. As a result

$$(E_i')^T = C\ B\ A\ (E_i)^T = C\ B\ H\ C^{-1}\ (E_i)^T$$

So that, if C is non-singular

$$C^{-1}\ C\ (E_i')^T = B\ H\ (E_i)^T$$

Since $C^{-1}\ C = I$, $B\ H = I$ for exact reproduction.

That is, $B = H^{-1}$.

Note: Curiously in this case the system can be designed without any reference to the spectral response curves of the observer! That is, B can be found from H, which does not depend on the observer's visual system in any way. The reason this is possible is that this system can actually duplicate the exact spectral distributions of intensity at each image point, since the light-sources it uses are just the same as those used to make the input image

If H happens to be diagonal, that is, if each sensor is carefully designed to pick up only one of the image generator inputs, then B can be diagonal. This usually can be achieved only with rather narrow sensitive band-widths and if there are regions of the spectrum where only one light-source contributes.

All these conditions are quite restrictive and unlikely to be met in practice, yet this corresponds to a technique used quite commonly for reproduction of photographic transparencies, where the film carrying the new image is exposed successively through three narrow-band filters with the old image.

A SYSTEM WHICH USES CONTROL OF DYE CONCENTRATIONS FOR REPRODUCTION

We next turn to a somewhat more complicated (and non-linear) case where three layers are superimposed, each with a different absorbing dye whose concentration can be controlled on a point-by-point basis (see fig. 5). _Photographic transparencies_ certainly fit this description. If we let $p_i(\lambda)$ be the transmission of unit concentrations of one of the three dyes, and C_i the actual concentration of this dye at a point, then the overall transmission $T(\lambda)$ of the sandwich is

$$T(\lambda) = \prod_{j=i}^{3} [p_j(\lambda)]^{C_j}$$

Sometimes it is more convenient to calculate the density instead, where the density is the logarithm of the inverse of the transmission.

$$D(\lambda) = \sum_{j=1}^{3} C_j \log_{10}[1/p_j(\lambda)]$$

Thus the overall _density_ is a linear combination of the densities of the individual dyes at unit concentration. This however helps little when one is calculating the stimulation levels in the observer's visual system when (s)he views the transparency using a light-source with spectral distribution $l(\lambda)$

$$E_j' = \int_{\lambda_0}^{\lambda_1} \sum_{j=i}^{3} [p_j(\lambda)]^{C_j} \, l(\lambda) \, e_i(\lambda) \, d\lambda$$

The stimulation levels are clearly related to the dye concentrations in a quite non-linear fashion, and the inner-product notation introduced earlier is of little help in analyzing this situation. While we can still produce a three-dimensional range of stimulation levels, it is difficult to determine without some computation what concentration levels $(C_i)^T$ are required to achieve a particular observer stimulation, $(E_i)^T$.

AN IDEALIZED MODEL FOR PHOTOGRAPHIC TRANSPARENCIES

In order to get some ideas of how to pick the dyes and how to control them, one can select conditions which will linearize the model to the point where previous methods for producing input controls to the image generation system apply. To do this, the following constraints must be applied:

1. To avoid multiplicative interactions between the three layers, at most one dye should absorb at a given wavelength. That is, if $p_i(\lambda) < 1$, then $p_j(\lambda) = 1$ for $i \neq j$.
2. In order to be able to produce "black" or very low transmission at all wavelengths, at least one dye should absorb at a given wavelength. That is, for every λ, at least one of the $p_i(\lambda)$ is less than 1.

These two conditions together imply that exactly one dye will absorb at a given wavelength. We can consequently divide the visible region of the spectrum into three sets, Λ_1, Λ_2 and Λ_3, such that $p_1(\lambda) < 1$, for $\lambda \varepsilon \Lambda_1$, while $p_2(\lambda) < 1$ for $\lambda \varepsilon \Lambda_2$ and $p_3(\lambda) < 1$ for $\lambda \varepsilon \Lambda_3$.

3. To ensure that each dye affects each of the three image sensors in constant proportion independent of concentration, the transmission should be a con-

stand less than one, for those wavelengths where it is not equal to one. Let this value be p_{io} for unit concentration of the ith dye (see fig. 6).

4. Finally the inputs to the image generator system must be transformed logarithmically to achieve <u>linear</u> control. That is, let $C_i = \log(P_i)/\log(p_{io})$.

We can now calculate the transmission of the sandwich. If $\lambda \varepsilon \Lambda_i$, then

$$T(\lambda) = [p_{io}]^{\log(P_i)/\log(p_{io})} = P_i$$

It is now convenient to split $l(\lambda)$ into three functions $l_1(\lambda)$, $l_2(\lambda)$ and $l_3(\lambda)$, where $l_i(\lambda) = l(\lambda)$ if $\lambda \varepsilon \Lambda_i$ and $l_i(\lambda) = 0$ otherwise. Then

$$l(\lambda) = \sum_{j=1}^{3} \ell_j(\lambda).$$

The stimulation levels can be calculated as follows:

$$E_i' = \int_{\lambda_0}^{\lambda_1} T(\lambda)\, l(\lambda)\, e_i(\lambda)\, d\lambda$$

$$E_i' = \int_{\lambda_0}^{\lambda_1} \sum_{j=i}^{3} P_j l_j(\lambda)\, e_i(\lambda)\, d\lambda$$

$$E_i' = \sum_{j=1}^{3} \underset{\sim}{l}_j \cdot \underset{\sim}{e}_i P_j$$

That is,

$$(E_i')^T = C\ (P_i)^T$$

where, the matrix $C = (c_{ij})$ and $c_{ij} = \underset{\sim}{e}_i \cdot \underset{\sim}{l}_j$. Note that $\underset{\sim}{l}_i$ here corresponds to P_i in the model that was analyzed earlier. Finally, the model has been idealized to the point where our previous methods apply directly. This has been achieved mostly by hypothesizing rather special dye-transmission curves which de-couple and linearize the system.

It should be noted that in practice dyes definitely do not obey the above mentioned restrictions and that as a result one ought to use the more precise model if accurate color reproduction is the goal. What is more, photographic film has further deficiencies which invalidate even the idealized model analysis. First of all, each sensitive layer is directly coupled to a dye layer, and no provision is made for cross-coupling as required in implementing the linear transform matrix B. Secondly, the spectral sensitivity curves of the photo-sensitive chemicals are not linear transforms of the human spectral response curves. Thirdly, the dye densities are not linearly related to image intensities -- in fact the reproduction invariably has higher contrast than the original. It is perhaps a little astonishing that one nevertheless finds color transparencies very pleasing!

DUPLICATING PHOTOGRAPHIC TRANSPARENCIES

After the slightly pessimistic results of the previous section it is perhaps worthwhile to point out that once again duplication can be performed with fair fidelity despite all the difficulties in reproducing arbitrary colored images. That is, despite the peculiar changes in the transmission of the layered film with changes in the concentrations of individual dyes, it is quite straightforward to determine the concentration of individual dyes. This follows from the linearity of the equation for density. Assume that one may use mono-chromatic light-sources of wavelength λ_1, λ_2 and λ_3 to sample the film, then one obtains a number of measurements of film density,

$$D_j = \sum_{i=1}^{3} C_i \log_{10}[1/p_i(\lambda_j)]$$

If we define a matrix $T = (t_{ij})$, where $t_{ij} = \log_{10}[1/p_i(\lambda_j)]$, then

$$(D_i)^T = T(C_i)^T$$

If this matrix T is non-singular we can obtain the concentrations quite easily from the measured densities at the three test wavelengths, using T^{-1}. Note that we effectively use image sensors with very narrow band sensitivities, quite unlike the general case, where we are forced to look for image sensors whose spectral response curves are linear transforms of the human spectral response curves.

By choosing the dyes carefully, it may further be possible to arrange for the matrix T to be diagonal by proper selection of the test wavelengths. That is, concentrations of one dye can be determined directly using density measurements at those wavelengths at which the other two dyes are transparent.

Further, it may be possible to arrange for the sensitive chemicals in the film which is to carry the reproduction to be separately sensitive to the three test wavelength. In this case accurate duplication can be achieved (within the limits of non-linearity and non-repeatability of photographic materials) simply by exposing the film with an image of the original successively through three narrow-band filters.

A SYSTEM WHICH USES CONTROL OF THE FRACTIONAL AREA COVERED BY INKS

Instead of controlling the concentration of the dyes on a point-by-point basis as in photographic methods, we may use dyes or inks of fixed concentration and instead vary the fraction of the surface covered with each ink. This may be attained by varying the dot size of ink-dots spaced in a regular pattern. This of course is the method used in lithographic reproduction of colored material. First consider a single ink. Assume that the transmission of the ink is $p_i(\lambda)$ and that a fraction A_i of the area is covered with the ink. If this dot-pattern has been applied to a substrate of reflectance $R_0(\lambda)$, the average reflectance will be

$$R(\lambda) = R_0(\lambda) \, [(1 - A_i) + A_i \, p_i(\lambda)]$$

This is so since a fraction A_i of the surface is covered with ink, while a fraction $(1 - A_i)$ is bare. The dots are usually spaced such that they are near the limits of resolution at normal viewing distance, and the dots corresponding to different inks lie on rasters which are rotated relative to one another to avoid the appearance of repeating patterns. The result is that dots of different inks overlap in different ways in various regions of print. Consequently, one may calculate the average reflectance of the completed print by multiplying the substrate reflectance and the transmission of each of the ink layers.

$$R(\lambda) = R_0(\lambda) \prod_{j=1}^{3} [(1 - A_j) + A_j p_j(\lambda)]$$

An alternate way of arriving at the same result is based on a calculation of the fractional areas covered by none of the inks, each of the inks in turn, two inks and finally all three inks (see fig. 7). We can now proceed to calculate the stimulation levels, given that the print is illuminated by a light-source with spectral distribution $l(\lambda)$,

$$E_i' = \int_{\lambda_0}^{\lambda_1} \prod_{j=i}^{3} [(1 - A_j) + A_j p_j(\lambda)] \, R_0(\lambda) l(\lambda) \, e_i(\lambda) \, d\lambda$$

Once again it is clear that the stimulation levels are related to the fractional area coverage factors in a non-linear fashion. We can certainly produce a three-dimensional range of stimulation levels, but it is non-trivial to determine what fractional area coverage values, $(A_i)^T$, will produce a particular set of stimulation levels, $(E_i)^T$.

AN IDEALIZED MODEL FOR LITHOGRAPHIC REPRODUCTION

Here again we may approximate the non-linear model to get some ideas on appropriate choices for the inks and methods of control. To do this, the following constraints must be applied:

1. To avoid interactions between the effects of the three printers, at most one ink should absorb at a given wavelength.
2. In order to be able to produce "black" or very low reflectance at all wavelengths, at least one ink should have zero transmission at a given wavelength.

These two conditions imply that exactly one ink will absorb at a given wavelength, and that it will absorb completely. Once again we can divide the visible spectrum into three regions Λ_1, Λ_2 and Λ_3, such that one dye absorbs in each region (see fig. 8).

3. Finally, the inputs to the image generator are complemented, that is, let $A_i = 1 - P_i$.

Then, if $\lambda \varepsilon \Lambda_i$,

$$R(\lambda) = R_0(\lambda) P_i$$

It is convenient again to define a set of functions $l_i(\lambda)$, in this case equal to $R_0(\lambda)\ l(\lambda)$ if $\lambda \varepsilon \Lambda_i$, and equal to zero otherwise. Then,

$$R_o(\lambda)\ l(\lambda) = \sum_{j=i}^{3} l_j(\lambda)$$

The calculation of the stimulation levels proceeds as follows,

$$E_i' = \int_{\lambda_0}^{\lambda_1} R(\lambda)\ l(\lambda)\ e_i(\lambda)\ d\ \lambda$$

$$E_i' = \int_{\lambda_0}^{\lambda_1} \sum_{j=i}^{3} P_j l_j(\lambda)\ e_i(\lambda)\ d\ \lambda$$

$$E_i' = \sum_{j=i}^{3} \underset{\sim}{l}_j \cdot \underset{\sim}{e}_i P_j$$

That is,

$$(E_i')^T = C(P_i)^T$$

where $C = (c_{ij})$ and $c_{ij} = \underset{\sim}{e}_i \cdot \underset{\sim}{l}_j$.

Note that again the spectral distribution $\underset{\sim}{l}_i$ corresponds to the light-source spectral output P_i in the first model that was analyzed.

Finally then, the model has been idealized to the point where our previous analysis applies. In order to do this, rather drastic assumptions had to be made regarding dye transmission functions. For exact reproduction the more precise model shown earlier must be used instead.

COLOR SEPARATION PHOTOGRAPHY AND MASKING

Modern lithographic reproductions of colored originals are of remarkably high quality. An important factor in achieving this high quality has been the realization that color separations photographed through three different filters should not be used directly to produce the offset plates. Instead each plate is made from a combination of the separations by a technique called "masking" [ ].

The most commonly used method depends on the superposition or "masking" of the film with a negative made by exposure through one filter, while the film is being exposed through another filter. Ignoring the nonlinearities of the photographic process, this corresponds to subtracting a fraction of the image made through one filter from another. By controlling exposure times and thus film densities, various amounts of "subtraction" can be achieved. Each final plate is as a result (approximately) a linear combination of the original images obtained through the three filters. That is, the dot-sizes at corresponding points in the three plates represent (approximately) a linear transform of the image intensities of the three filtered images.

This linear transform corresponds to the matrix $B = (AC)^{-1}$, which is needed between the image sensor outputs and the image generator inputs in the idealized linear model. Masking thus accounts for the off-diagonal terms in the matrix B, which in turn are a function of the filter curves, the spectral response curves of the human observer and the (idealized) ink transmission curves. It is often (falsely) stated that masking is required to deal with imperfections in the ink transmission curves -- whereas it has just been shown that masking is required with "ideal" inks.

Imperfections in the ink, in terms of departures from the ideal model presented in the previous section, are not taken care of by masking. As a result of the non-linearity of the general case, reproduction can only be exact for a small number of ink combinations, and will be approximate for others. In fact, the masking variables (the exposure time required for each "mask") are usually determined empirically by using a standard original with several color patches. The exposures are adjusted until these are reproduced correctly. The color patches usually include the three printing inks, three patches in which two inks are superimposed and three or so "neutral" colors (white, gray and black).

Note, that as a result, the system is actually tuned for duplication. As was shown earlier, this is satisfactory for reproduction of arbitrary originals only if the input sensitivity curves are linear transforms of the human spectral response curves. In some cases it may be satisfactory to tune the system instead to reproduction of a particular kind of input material, for example, a particular make of photographic film [ ].

This discussion of masking has of necessity been over-simplified and has ignored such techniques as highlight pre-masking and unsharp masking, techniques which help overcome the non-linearities and dynamic range limitations of the process.

COMPUTATIONAL METHODS

For each of the models of image reproduction systems presented, an expression was exhibited for the stimulation of the observer's visual system as a function of the control inputs to the image generators. The three expressions were:

$$E_i' = \sum_{j=1}^{3} P_j \int_{\lambda_0}^{\lambda_1} p_j(\lambda)\, e_i(\lambda)\, d\lambda$$

$$E_i' = \int_{\lambda_0}^{\lambda_1} \prod_{j=1}^{3} [p_j(\lambda)]^{C_j}\, l(\lambda)\, e_i(\lambda)\, d\lambda$$

$$E_i' = \int_{\lambda_0}^{\lambda_1} \prod_{j=1}^{3} [(1 - A_j) + A_j p_j(\lambda)]\, R_0(\lambda)\, l(\lambda)\, e_i(\lambda)\, d\lambda$$

The control inputs are $(P_i)^T$ -- the light-source intensities, $(C_i)^T$ -- the dye concentrations, and $(A_i)^T$ -- the fractional areas covered by ink. Only in the first case is it possible to solve directly for the control inputs given the desired observer stimulation levels $(E_i)^T$. In the other cases, one has to resort to trial-and-error or hill-climbing search techniques, unless one chooses to accept inaccuracies in order to linearize the model. It would be very inefficient to do this computation afresh every time a new point in the image is analyzed.

Accordingly, one may imagine performing this calculation ahead of time and recording the results in some sort of look-up table. This implies however that only a finite number of possible stimulation levels can be explored and then replicated. That is, the three-dimensions have to be quantized suitably. This may be a problem in a high quality system, since one might have to divide the intervals quite finely and the look-up table may become unwieldy. If for example it is found that the quality of the process is such that a hundred divisions are needed along each of the three dimensions, then the total look-up table would contain a million entries. Each entry is a set of three numbers to be used as settings for the control inputs to the image generators. The table is entered at a location that corresponds to the desired observer stimulation levels (see fig. 9). This table is quite large, and even with modern day storage methods may be too costly.

At the expense of slightly increased computation, the table can be substantially shrunk by using a much coarser quantization and simple interpolation between entries. If, for example, this allows one to divide each dimension into only ten intervals, then the whole table contains only three times a thousand numbers and can easily be accomodated in a small read-only-memory (ROM) module for example. Electronics must however then be added to perform the interpolation, and eight table entries are accessed for every look-up operation.

If the table is organized with the observer stimulation as its axes, it will be quite general and work with any image sensing system, as long as the linear transform A^{-1} is first applied to the outputs of the image sensors. Alternatively, the table can be organized directly with the image sensor outputs as axes. Since not all combinations of observer stimulation levels are possible, regions of the table will be blank. Similarly, not all combinations of image sensor stimulations will be possible. One may be able to achieve some storage economy by compacting the table accordingly, or else using these areas of the memory to store other information.

The entries in the table may be filled in as indicated above by calculations based on the models. In practice, there are likely to be discrepancies between the model and reality and it may be helpful to determine some of the points empirically. In particular, it is possible that the exact dye absorption curves are not known, or that interactions among inks take place that are not modelled. It is however impractical to fill in the whole table in this way. Techniques may be used for interpolating between empirically determined table entries using the structure supplied by the equations derived from the model.

SUMMARY AND CONCLUSIONS

For accurate reproduction of arbitrary colored images, the image sensor's spectral response curves must be linear transforms of the spectral response curves of the human visual system. Aside from this general constraint, a system for the reproduction of colored images must be designed in such a way that the image generators produce the appropriate excitations in the receptors of the observer's visual system. Three specific systems were analyzed and the necessary computation of image generator inputs from image sensor outputs was detailed.

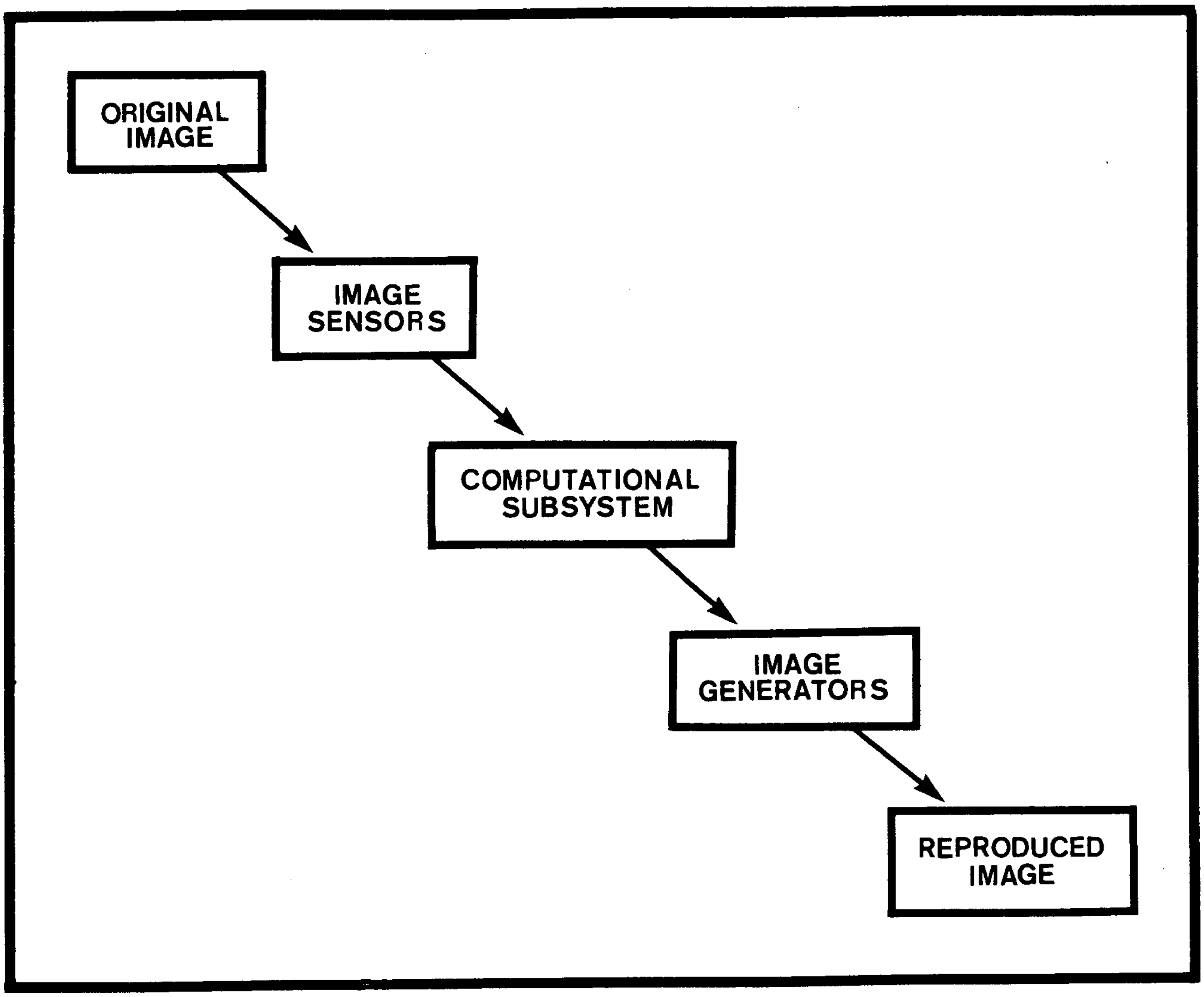
In only one case, color television, could this computation be accomplished analytically, and in this case it turned out to be a simple linear transformation. In the other cases studied, photographic transparencies and lithographic printing, the computations were straightforward only when simplifications were introduced in the form of unrealizable dye absorption curves. It was suggested that point-by-point computation of image generator inputs is now feasible and this is in fact the only way to achieve accurate reproduction with practical dyes or inks. The well known color separation and masking operation is seen to be only the linear transform which applies when inks with idealized absorption curves are used, and does not deal with "ink-imperfections".

Other topics dealt with, include the proper point for adjustments to compensate for observer adaptation, the use of more than three image generators, and the duplication of colored images. Such inexact notions as "primary color", "secondary color" and "complementary color" were studiously avoided.

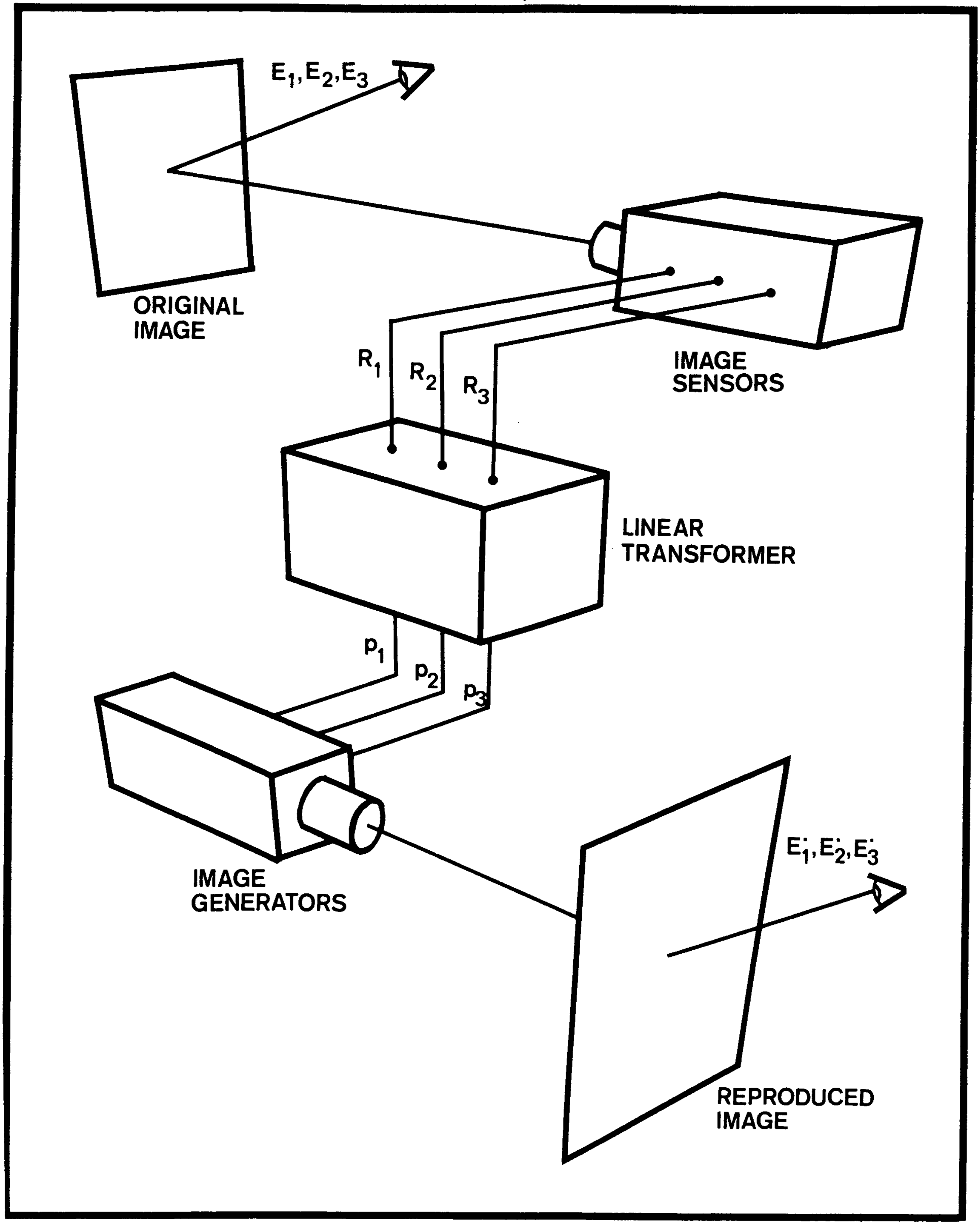
These techniques will be of immediate importance where images are already scanned and transmitted, since the simple table look-up computation developed here can be easily incorporated in such a system. These methods will also be of importance when colors are to be judged in images which are transmitted from locations inaccessible to (wo)man, such as other planets.

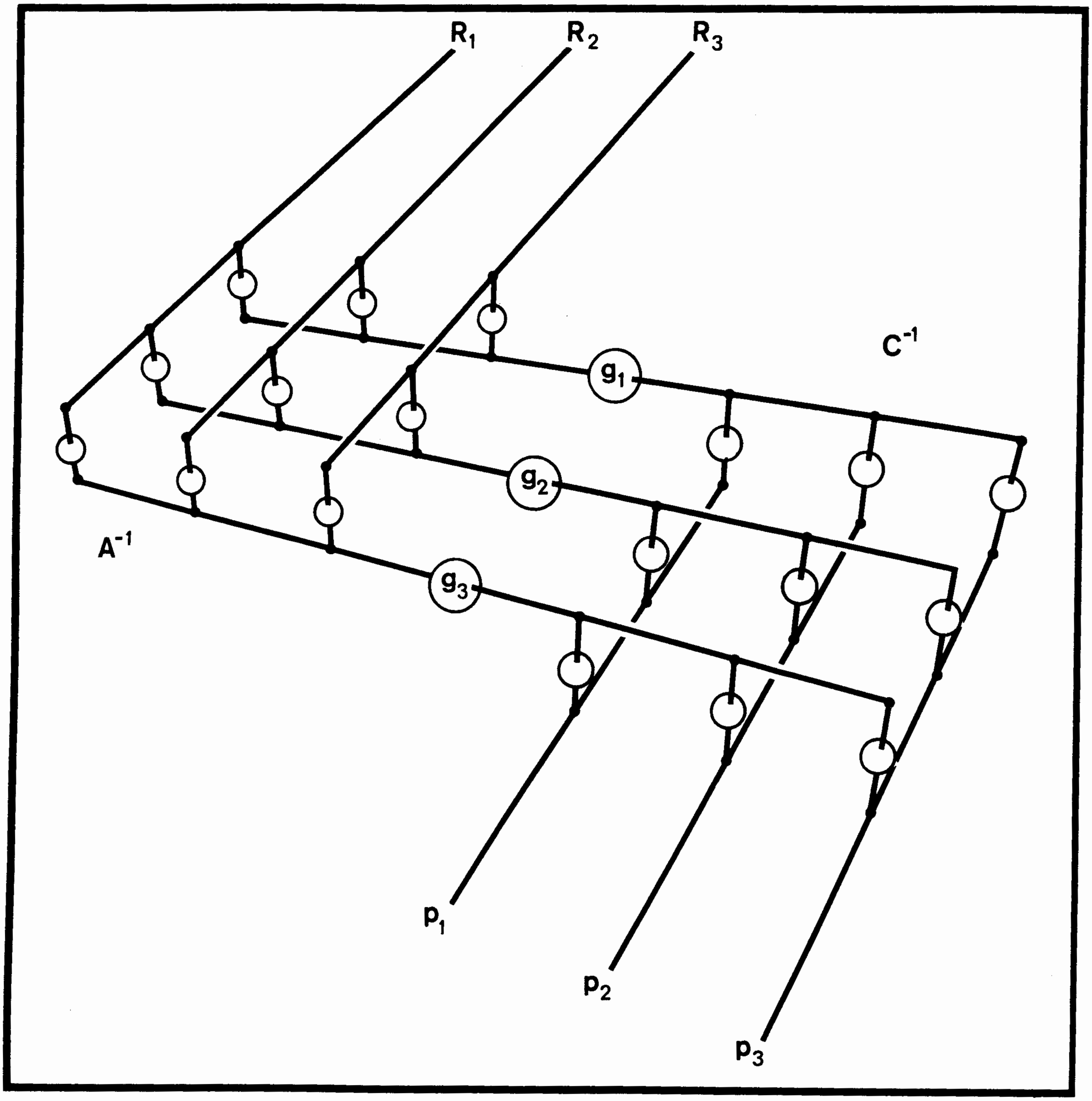
ACKNOWLEDGEMENTS

I would like to thank William Silver and Robert Buckley for a number of helpful discussions.

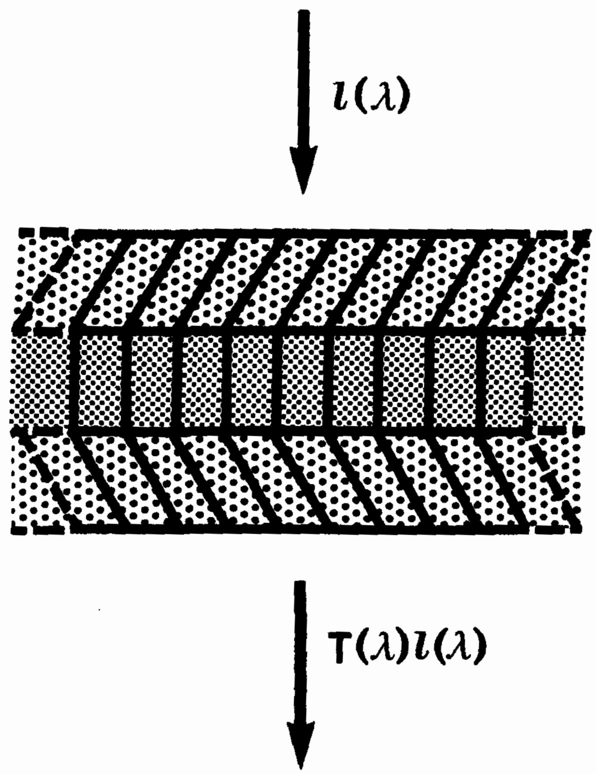

ORIGINAL IMAGE
IMAGE SENSORS
COMPUTATIONAL SUBSYSTEM
IMAGE GENERATORS
REPRODUCED IMAGE

	LIGHT SOURCE	TRANSMISSION ABSORPTION	REFLECTION ABSORPTION
ADDITIVE	COLOR TELEVISION LASER PROJECTOR	MAXWELL & IVES SUPERIMPOSED PROJECTORS	JOLY'S STRIP FILTERS LUMIERE'S AUTOCHROME
AREA CONTROL MULTIPLICATIVE	NOT APPLICABLE	TRANSMISSION LITHOGRAPHY	LITHOGRAPHY
CONCENTRATION CONTROL MULTIPLICATIVE	NOT APPLICABLE	PHOTOGRAPHIC TRANSPARENCIES	PHOTOGRAPHIC PRINTS CARBRO PRINTS DYE TRANSFERS COLOR XEROGRAPHY

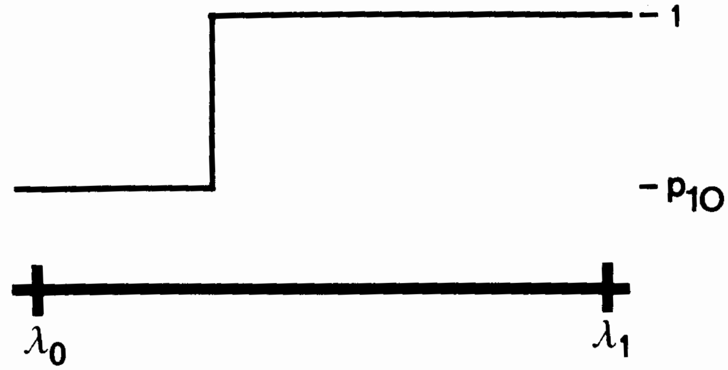

E1, E2, E3
ORIGINAL IMAGE
IMAGE SENSORS
R1
R2
R3
LINEAR TRANSFORMER
p1
p2
p3
IMAGE GENERATORS
E1', E2', E3'
REPRODUCED IMAGE



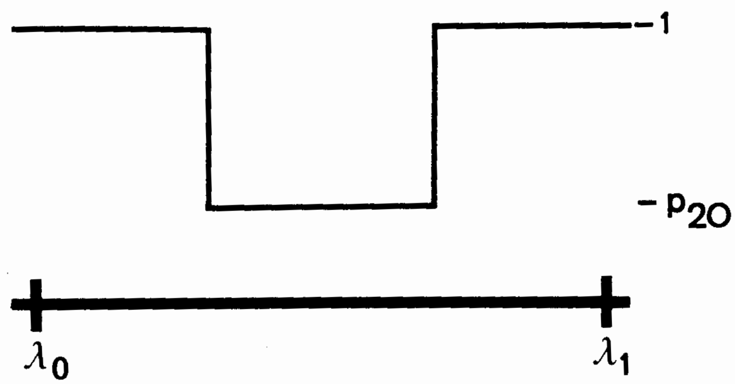
R1
R2
R3
C-1
g1
g2
g3
A-1
p1
p2
p3



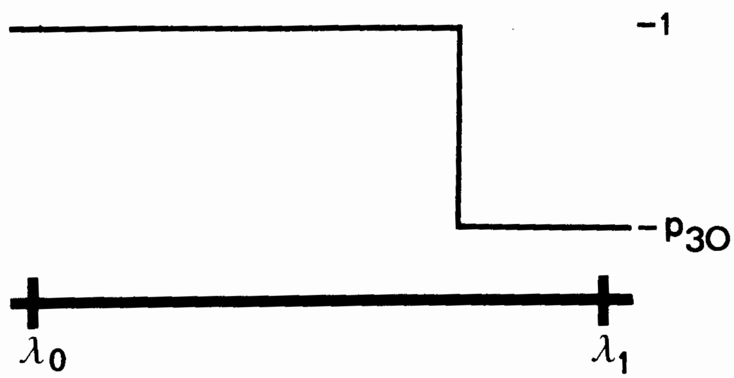
Dye-transmission	Concentration
$p_1(\lambda)$	C_1
$p_2(\lambda)$	C_2
$p_3(\lambda)$	C_3



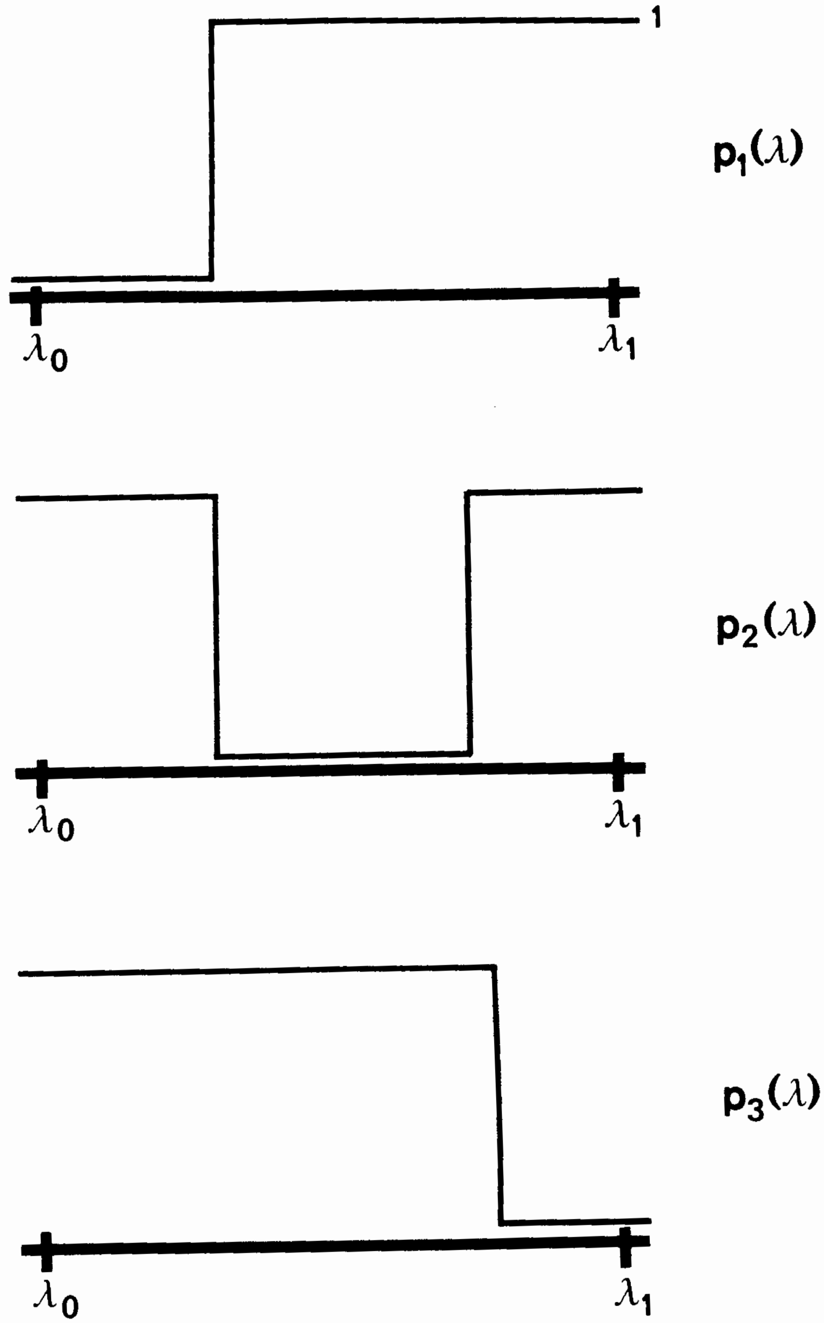
$p_1(\lambda)$

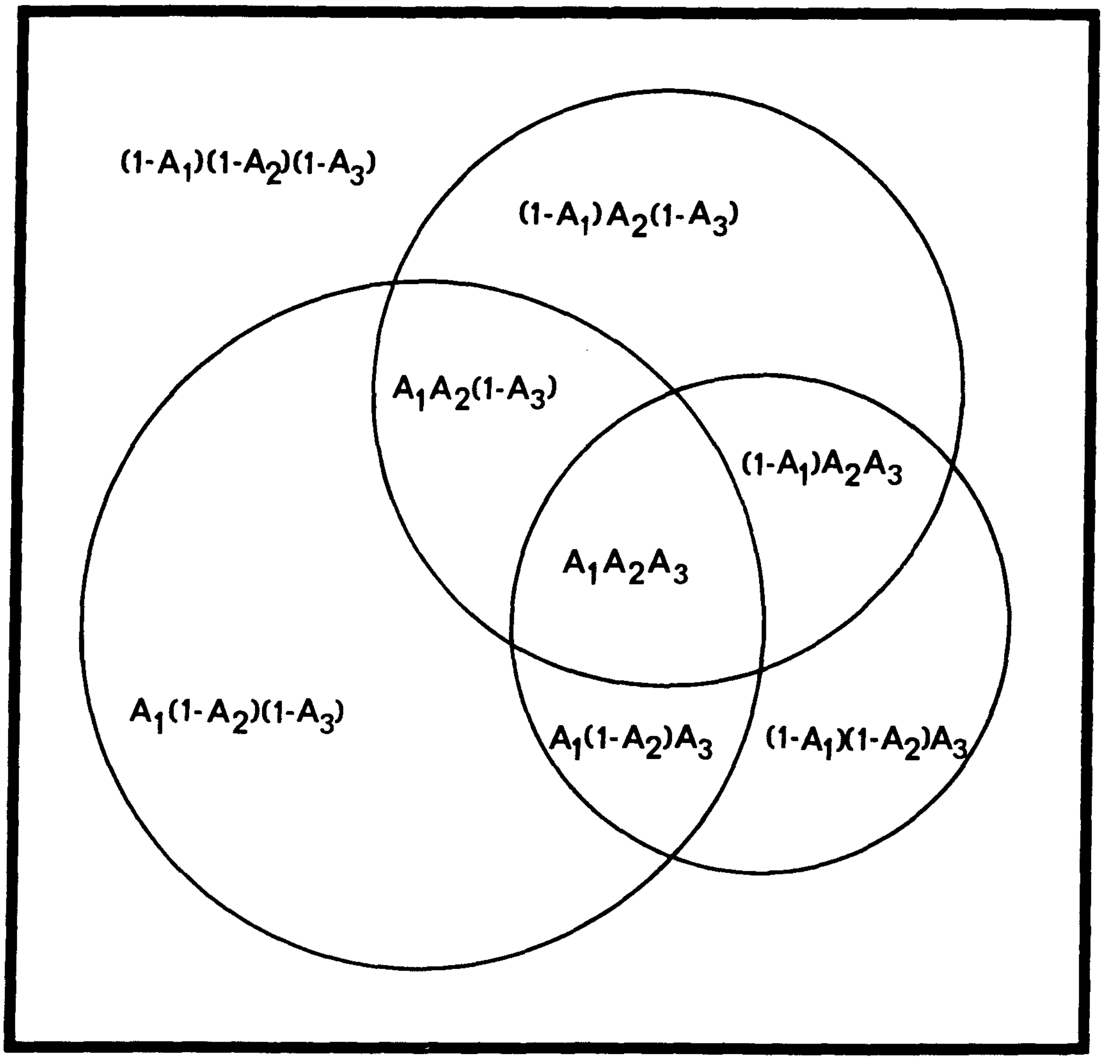


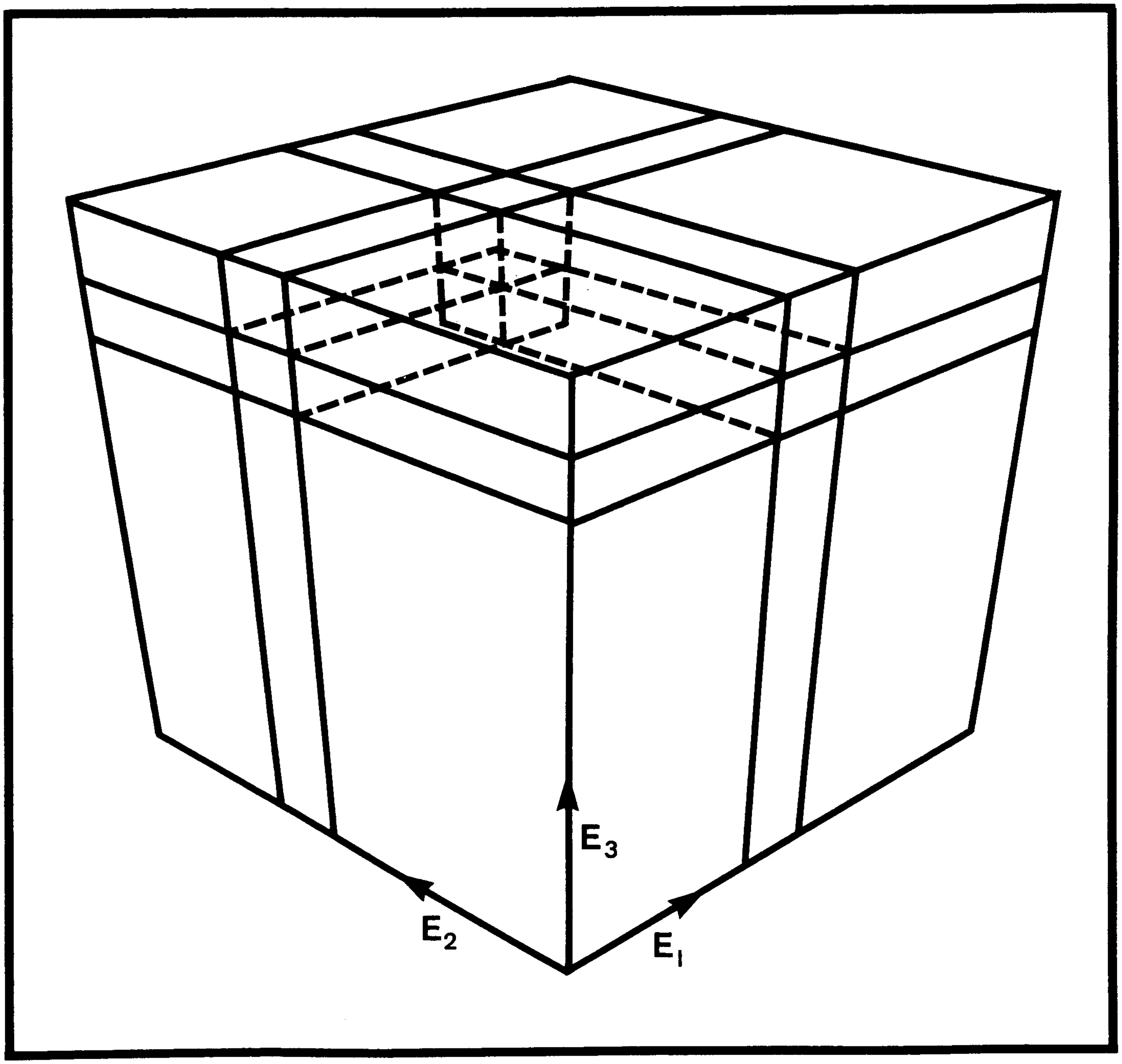
$p_2(\lambda)$



$p_3(\lambda)$


1
$p_1(\lambda)$
λ_0
λ_1
$p_2(\lambda)$
λ_0
λ_1
$p_3(\lambda)$
λ_0
λ_1


(1-A1)(1-A2)(1-A3)
(1-A1)A2(1-A3)
A1A2(1-A3)
(1-A1)A2A3
A1A2A3
A1(1-A2)(1-A3)
A1(1-A2)A3
(1-A1)(1-A2)A3


E3
E2
E1

REFERENCES:

COLOR REPRODUCTION TECHNIQUES

COLOR SCIENCE AND COLORIMETRY

COLOR PERCEPTION -- HISTORIC

COLOR PERCEPTION -- CURRENT

PSYCHOPHYSICS & NEURO-PHYSIOLOGY

VECTOR SPACES AND INNER PRODUCTS

LINEAR PROGRAMMING AND PSEUDO-INVERSES

Color Perception -- Current:

1. G. S. Brindley, Physiology of the retina and the visual pathway, Edward Arnold Ltd., London (1970), 199-259.

2. T. N. Cornsweet, Visual Perception, Academic Press, New York (1970), 155-267.

3. R. Evans, The Perception of Color, John Wiley & Sons, New York (1974).

4. R. P. Feynman, R. B. Leighton and M. Sands, The Feynman Lectures on Physics, Addison-Wesley, Reading, MA (1963), Chapters 35 and 36.

5. C. H. Graham (ed.), Vision and Visual Perception, Wiley, New York (1965).

6. R. N. Haber and M. Hershenson, The Psychology of Visual Perception, Holt, Rinehart & Winston, New York (1973), 60-85.

7. Le Grand, Light, Colour, Vision, Chapman & Hall, London (1967).

8. F. W. Sears, Optics, Addison-Wesley, Reading, MA (1958), 350-369.

9. Symposium No. 8, Visual Problems of Colour, Vol. I and II, National Physics Laboratory, Teddington, England, September 1957; Her Majesty's Stationary Office, London (1958).

Psychophysics and Neurophysiology:

1. H. Autrum and I. Thomas, "Comparative Physiology of Colour Vision in Animals," in Handbook of Sensory Physiology, Vol. VII/3 Part A, Springer Verlag, New York (1973), 661-692.

2. W. L. Brewer, "Fundamental response functions and binocular color matching," J. Opt. Soc. Am., 44, 207-212 (1954).

3. R. L. DeValois , "Analysis and coding of colour vision in the primate visual system," Cold Spring Harbor Symposia on Quantitative Biology, 30, 567-579 (1965).

4. R. L. DeValois, "Central mechanisms of color vision," Handbook of Sensory Physiology, Vol. VII/3 Part A, Springer-Verlag, New York (1973), 209-253.

5. H. Helson, "Fundamental problems in color vision, I," J. Exper. Psychol., 23 (1938).

6. H. Helson, "Fundamental problems in color vision, II," J. Exper. Psychol., 26 (1940).

7. L. M. Hurvich and D. Jameson, "An opponent-process theory of color vision," Psychol. Rev., 64, 384-404 (1957).

8. D. Judd, "Hue, saturation and lightness of surface colors with chromatic illumination," J. Opt. Soc. Am., 30, 2-32 (1940).

9. E. H. Land, "Experiments in color vision," Sci. Am., 200, 84-89 (1959).

10. J. Y. Lettvin, "The color of colored things," Quarterly Progress Report, 87, Research Laboratory of Electronics, MIT (1967).

11. D. M. Purdy, "The Bezold-Bruecke phenomena and contours for constant hue," Amer. J. Psychol., 49, 313-315 (1937).

12. W. Richards and E. Parks, "Model for color conversion," J. Opt. Soc. Am., 61, 971-976 (1971).

13. W. A. G. Rushton, "Visual pigments in man," Sci. Am., 207, 120-132 (19).

14. G. Wald, P. K. Brown and I. R. Gibbons, "The problem of visual excitation," J. Opt. Soc. Am, 53, 20-35 (1963).

15. G. Wlad, "The receptors for human color vision," Science, 145, 1007-1017 (1964).

16. G. Wald and P. Brown, "Human Color Vision and Color Blindness," Cold Spring Harbor Symposia on Quantitative Biology, 30, 567-579 (1965).

17. S. M. Zeki, "Colour coding in the rhesus monkey pre-striate cortex," Brain Research, 53, 422-427 (1973).

Vector spaces:

1. I. N. Herstein, Topics in Algebra, Blaisdell Publishing Co., Waltham, MA (1964), 130-143, 150-159, 216-225.

2. G. Birkhoff and S. MacLane, A Survey of Modern Algebra, Mac-Millan Co., New York (1965), 149-179, 189-200.

3. F. B. Hildebrand, Methods of Applied Mathematics, 2nd edition, Prentice-Hall, Inc., Englewood Cliffs, N.J. (1965) 4-9, 23-30, 81-88.

4. J. T. Moore, Elements of Linear Algebra and Matrix Theory, McGraw-Hill, New York (1968), 63-139, 178-235.

5. I. E. Segal and R. A. Kunze, Integrals and Operators, McGraw-Hill, New York (1968), 123-174, 221-232.

Linear programming and pseudo-inverses:

1. A. Ben-Israel and T. N. Greville, Generalized Inverses - Theory and Applications, Wiley-Interscience, New York (1974).

2. T. L. Boullion and P. L. Odell, Generalized Inverse Matrices, Wiley-Interscience, New York (1971).

3. F. Hillier and G. J. Lieberman, Introduction to Operations Research, Holden-Day, San Francisco (1967).

4. N. P. Loomba, Linear Programming -- An Introductory Analysis, McGraw-Hill, New York (1964).

5. D. G. Luenberger, Introduction to Linear and Non-linear Programming, Addison-Wesley, Reading, MA (1965).

Color photography:

1. Henney and Dudley, Handbook of Photography, McGraw-Hill, New York (1939).

2. C. E. Mees and J. H. James, The Theory of the Photographic Process, 3rd edition, MacMillan Company, New York (1966).

3. C. B. Neblette, Photography, Its Materials and Processes, Van Nostrand Company, New York (1952).

4. H. E. J. Neugebauer, "Quality factor for filters whose spectral transmittances are different from color mixture curves, and its application to color photography," Journal of the Optical Society of America, 46,

APPENDIX. Review of some relevant properties of infinite dimensional vector spaces.

Definition: Let V be a vector-space over a field F. Then the orthogonal complement $A^{\perp}$ of a subspace A of V is the set of vectors perpendicular to all vectors in A. That is,

$$A^{\perp} = \{\underset{\sim}{x} \mid \underset{\sim}{x} \cdot \underset{\sim}{a} = 0 \text{ for all } \underset{\sim}{a} \varepsilon A\}$$

Lemma 1: If A is a finite-dimensional subspace of dimension n with basis $\{a_i\}$ in V, then the orthogonal complement $A^{\perp}$ is the set of vectors perpendicular to all basis vectors. That is,

$$A^{\perp} = \{\underset{\sim}{x} \mid \underset{\sim}{x} \cdot \underset{\sim}{a}_i = 0 \text{ for } i = 1 \text{ to } n\}$$

Proof: If A is finite dimensional, every vector $\underset{\sim}{a} \varepsilon A$ can be expressed as a sum of scaled basis vectors as follows,

$$\underset{\sim}{a} = \sum_{i=1}^{n} \alpha_i \underset{\sim}{a}_i$$

where the α_i are in the field F. A vector $\underset{\sim}{x}$ perpendicular to each of the $\underset{\sim}{a}_i$ will clearly be perpendicular to any $a \varepsilon A$. Conversely, a vector $\underset{\sim}{x}$ perpendicular to each $a \varepsilon A$, will certainly also be perpendicular to each of the basis vectors $\underset{\sim}{a}_i$.

Lemma 2: If A and B are subspaces of the vector-space V and $A \supset B$, then $A^{\perp} \subset B^{\perp}$.

Proof: Consider a vector $\underset{\sim}{x} \varepsilon A^{\perp}$, then $\underset{\sim}{x} \cdot \underset{\sim}{a} = 0$ for all $\underset{\sim}{a} \varepsilon A$. Since $A \supset B$, this implies that $\underset{\sim}{x} \cdot \underset{\sim}{b} = 0$ for all $\underset{\sim}{b} \varepsilon \underset{\sim}{B}$ as well. Therefore $\underset{\sim}{x} \varepsilon B^{\perp}$. Consequently, $A^{\perp} \subset B^{\perp}$.

Definition: The perpendicular projection of a vector $\underset{\sim}{v} \varepsilon V$ on a finite dimensional subspace A is the vector $\underset{\sim}{a} \varepsilon A$, which is closest to $\underset{\sim}{v}$, that is, the vector which minimizes $(\underset{\sim}{v} - \underset{\sim}{a}) \cdot (\underset{\sim}{v} - \underset{\sim}{a})$.

Lemma 3: The perpendicular projection of a vector $\underset{\sim}{v} \varepsilon V$ on a finite dimensional subspace A is the vector $\underset{\sim}{a}$ given by

$$\underset{\sim}{a} = \sum_{i=1}^{n} \alpha_i \underset{\sim}{a}_i \quad \text{where } (\alpha_i)^T = M^{-1}(\underset{\sim}{v} \cdot \underset{\sim}{a}_i)^T$$

Here $\{\underset{\sim}{a}_i\}$ is a basis for A and $M = \{m_{ij}\}$ is the symmetric Gram or normal matrix, with $m_{ij} = \underset{\sim}{a}_i \cdot \underset{\sim}{a}_j$.

Proof: We wish to minimize $(\underset{\sim}{v} - \underset{\sim}{a}) \cdot (\underset{\sim}{v} - \underset{\sim}{a}) = \underset{\sim}{v} \cdot \underset{\sim}{v} - 2\underset{\sim}{a} \cdot \underset{\sim}{v} + \underset{\sim}{a} \cdot \underset{\sim}{a}$. That is,

$$\underset{\sim}{v} \cdot \underset{\sim}{v} - 2 \sum_{i=1}^{n} \alpha_i (\underset{\sim}{v} \cdot \underset{\sim}{a}_i) + \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j (\underset{\sim}{a}_i \cdot \underset{\sim}{a}_j)$$

That is, $(\underset{\sim}{v} \cdot \underset{\sim}{a}_i)^T = M(\alpha_i)^T$. The result follows, since the basis vectors are linearly independent and M is therefore nonsingular and consequently has an inverse M^{-1}.

Lemma 4: If $\underset{\sim}{a}$ is the perpendicular projection of the vector $\underset{\sim}{v} \varepsilon V$ on the finite dimensional subspace A, then $\underset{\sim}{x} = \underset{\sim}{v} - \underset{\sim}{a}$ is in $A^{\perp}$.

Proof: From the previous lemma we have

$$\underset{\sim}{v} \cdot \underset{\sim}{a}_k = \sum_{i=1}^{n} \alpha_i (\underset{\sim}{a}_i \cdot \underset{\sim}{a}_k) = \underset{\sim}{a} \cdot \underset{\sim}{a}_k$$

So, $(\underset{\sim}{v} - \underset{\sim}{a}) \cdot \underset{\sim}{a}_k = 0$ for k = 1 to n. So $\underset{\sim}{x} = (\underset{\sim}{v} - \underset{\sim}{a}) \varepsilon A^{\perp}$.

Lemma 5: If A is a finite dimensional subspace of dimension n of the vector-space V, then any vector $\underset{\sim}{v} \varepsilon V$ can be written as the sum of a vector $\underset{\sim}{a} \varepsilon A$ and a vector $\underset{\sim}{x} \varepsilon A^{\perp}$ (that is, $V = A \oplus A^{\perp}$, the direct sum of A and $A^{\perp}$).

Proof: This follows directly from the previous two lemmas if we let $\underset{\sim}{v} = \underset{\sim}{a} + \underset{\sim}{x}$, where $\underset{\sim}{a}$ is as defined in lemma 3 and $\underset{\sim}{x}$ as in lemma 4. (It is also clear that the decomposition into a vector in A and a vector in $A^{\perp}$ is unique).

Lemma 6: If $B = A^{\perp}$ is the orthogonal complement of a finite dimensional subspace A <u>and</u> $\underset{\sim}{x} \cdot \underset{\sim}{x} = 0$ implies $\underset{\sim}{x} = 0$ if $\underset{\sim}{x} \varepsilon B$, then the orthogonal complement of B is A. That is, $(A^{\perp})^{\perp} \quad A$.

Proof: Consider $\underset{\sim}{v} \varepsilon B^{\perp}$. By the previous lemma we can decompose $\underset{\sim}{v}$ into a sum $\underset{\sim}{a} + \underset{\sim}{x}$, where $\underset{\sim}{a} \varepsilon A$ and $\underset{\sim}{x} \varepsilon A^{\perp}$. Then,

$$\underset{\sim}{v} \cdot \underset{\sim}{x} = (\underset{\sim}{a} + \underset{\sim}{x}) \cdot \underset{\sim}{x} = \underset{\sim}{a} \cdot \underset{\sim}{x} + \underset{\sim}{x} \cdot \underset{\sim}{x}$$

Now $\underset{\sim}{v} \cdot \underset{\sim}{x} = 0$, since $\underset{\sim}{x} \varepsilon B$ and $\underset{\sim}{v} \varepsilon B^{\perp}$. Also $\underset{\sim}{a} \cdot \underset{\sim}{x} = 0$, since $\underset{\sim}{a} \varepsilon A$ and $\underset{\sim}{x} \varepsilon A^{\perp}$. Therefore $\underset{\sim}{x} \cdot \underset{\sim}{x} = 0$. By assumption this implies that $\underset{\sim}{x} = 0$. Therefore $\underset{\sim}{v} = \underset{\sim}{a}$, and so $\underset{\sim}{v} \varepsilon A$. Therefore $B^{\perp} = A$ or $(A^{\perp})^{\perp} = A$.

Lemma 7: If A and B are finite-dimensional subspaces of a vector-space V, of equal dimension n, say, and $A \subset B$, then $A = B$.

Proof: If A is finite dimensional, there must exist a set of n linearly independent vectors $\{a_i\}$, which span A. If $A \subset B$, then these same n vectors are also in B. Since any n linearly independent vectors in B will form a basis for the vector space B, these vectors will. That is, B is spanned by the vectors $\{a_i\}$. Therefore $A = B$.

Lemma 8. Let A be an n x m matrix with $n < m$, then the underdetermined set of equations $A \underset{\sim}{x}^T = \underset{\sim}{y}^T$ has a solution of minimum norm of the form,

$$\underset{\sim}{x}^T = (A^T A)^{-1} A^T \underset{\sim}{y}^T$$

if A^TA is non-singular. The norm being minimized is $\underset{\sim}{x}\,\underset{\sim}{x}^T$.

Proof: Introduce the Lagrangian multiplier λ say and minimize,

$$\underset{\sim}{x}\,\underset{\sim}{x}^T + \lambda(A\,\underset{\sim}{x}^T - \underset{\sim}{y}^T)^T (A\,\underset{\sim}{x}^T - \underset{\sim}{y}^T)$$

$$= \underset{\sim}{x}\,\underset{\sim}{x}^T + \lambda(\underset{\sim}{x}\,A^T - \underset{\sim}{y})(A\,\underset{\sim}{x}^T - \underset{\sim}{y}^T)$$

$$= \underset{\sim}{x}\,\underset{\sim}{x}^T + \lambda(\underset{\sim}{x}\,A^T A\,\underset{\sim}{x}^T - \underset{\sim}{y}\,A\,\underset{\sim}{x}^T - \underset{\sim}{x}\,A^T \underset{\sim}{y}^T + \underset{\sim}{y}\,\underset{\sim}{y}^T)$$

Differentiating with respect to $\underset{\sim}{x}$,

$$\underset{\sim}{x}^T + \underset{\sim}{x}^T + \lambda[(A^T A\,\underset{\sim}{x}^T) + (\underset{\sim}{x}\,A^T A)^T - (\underset{\sim}{y}A)^T - (A^T\underset{\sim}{y}^T) + 0]$$

Dividing by two and setting the result equal to zero,

$$\underset{\sim}{x}^T + \lambda(A^TA\underset{\sim}{x}^T - A^T\underset{\sim}{y}^T) = 0$$

If this is to hold as λ becomes very large, then,

$$A^T A\,\underset{\sim}{x}^T - A^T \underset{\sim}{y}^T = 0$$

If A^TA, a n x n matrix, is non-singular, we have,

$$\underset{\sim}{x}^T = (A^TA)^{-1} A^T \underset{\sim}{y}^T$$

The m x n matrix $(A^T A)^{-1} A^T$ is called the pseudo-inverse of A.