Computational Mechanisms for Colour and Lightness Constancy

Wayne Holmes Caplinger

Ph.D. Thesis

University of Edinburgh

1984

Computational Mechanisms for Colour and Lightness Constancy

Abstract

Attributes of colour images have been found which allow colour and lightness constancy to be computed without prior knowledge of the illumination, even in complex scenes with three-dimensional objects and multiple light sources of different colours. The ratio of surface reflectance colour can be immediately determined between any two image points, however distant. It is possible to determine the number of spectrally independent light sources, and to isolate the effect of each. Reflectance edges across which the illumination remains constant can be correctly identified.

In a colour image all the pixels of a single surface colour lie in a single structure in flux space. The dimension of the structure equals the number of illumination colours. The reflectance ratio between two regions is determined by the transformation between their structures. Parallel tracing of edge pairs in their respective structures identifies an edge of constant illumination, and gives the lightness ratio of each such edge. Enhanced noise reduction techniques for colour pictures follow from the natural constraints on the flux structures.

In a scene illuminated by multiple distant point sources of distinguishalbe colours, the spatial angle between the sources and their brightness ratios can be computed from the image alone. If there are three or more sources then reflectance constancy is immediately possible without use of additional knowledge.

The results are an extension of Edwin Land's Retinex algorithm. They account for previously unexplained data such as Gilchrist's

veiling luminances and his single-colour rooms.

The validity of the algorithms has been demonstrated by implementing them in a series of computer programs. The computational methods do not follow the edge or region finding paradigms of previous vision mechanisms. Although the new reflectance constancy cues occur in all normal scenes, it is likely that human vision makes use of only some of them.

Acknowledgments

I am indebted to Jim Howe, my supervisor, for introducing me to this problem, for his advice, and for many useful suggestions on the manuscript. Mathematical assistance came from discussions with Robert Fisher and John Hallam. Robert also co-authored with me a library of vision processing tools. Jo Cumming and Dave Lee commented on various drafts. Steven Cameron provided the graphics tools used in research and presentation. John Hallam and Roy Featherstone gave advice on formatting and printing. Jo Cumming and John Bateman's practical assistance made the deadline achievable. David Monk introduced me to Projective Spaces.

The difficulties were lightened by the encouragement and friendship of the other members of the Department of Artificial Intelligence, in particular my colleagues in the vision group: Aravinda Athukorala, Bob Beattie, Andrew Blake, Robert Fisher, John Hallam, Chris Malcolm, Barry Shephard, and Bao-Lin Yin.

This thesis was composed by me and describes my own work.

Contents

Chapter One

Description of Colour Constancy
Previous and New Algorithms and Domains

Traditionally, colour photographic film comes as either daylight, for use outdoors, or tungsten, for use indoors. If the tungsten film is used under sunlight the photograph has an overall blue tint. If the daylight film is used indoors, under incandescent bulbs, the photograph will have an orange wash. The reason for this is well known; the sun is a hotter black body radiator than the tungsten filament, and so the spectrum of its light is stronger towards the blue end. Thus outdoors there is more blue light reaching the objects, and on average more blue light being reflected and so available to the camera. The film manufacturers compensate for the colour of sunlight or tungsten light when designing the emulsions. All of this is well understood. However, it leaves the question of why a human observer does not see the orange or blue wash; it does not give a mechanism by which the human eye and brain can compensate for different illumination colours.

A common explanation is to invoke some "adaptive" mechanism. This term has been used for many different phenomena. One of the earliest was the adjustment of the eye to the overall level and colour of illumination. For example, suppose that the pigments of the colour receptors, the cones, are bleached by the light. If there is a lot of blue, they become proportionately less sensitive to blue, and similarly with other colours. Although this does occur, it can not wholly explain colour constancy. This form of retinal adaptation takes a noticeable amount of time, seconds or even minutes. If this

were the method of colour constancy, then the human eye would see a coloured wash every time it entered or exited a building, and this wash would last until adaptation was complete. This does not match common observation. A second problem is that the bleaching is local, so the portion of the retina observing a red object would become less sensitive to red, while a neighbouring piece would become less sensitive to whatever colour was reflected from its object.

Another version of the "adaptive" explanation is that some neurological mechanism between the eye and visual cortex determines the average flux in the scene and compensates for it across the scene. Irrespective of the mechanism proposed for finding the average flux, it has been shown that the eye does not compensate for the lighting colour by using the average flux colour from the scene. One experiment [McCann, Hall and Land, 1977] studied various scenes of differing average reflectances. A red bias to the average flux from a scene could be caused by red light, or by a scene with many or large reddish objects. As expected, human observers were not fooled by average fluxes caused by the objects in the scene; they compensated only when the bias was caused by the lighting. Two scenes of identical average flux are interpreted differently if their illuminations are different. This leaves the problem of how could the eye distinguish the effects of light and reflectance, when both influence the received flux equally and universally.

Several other demonstrations can be devised to show that such simple mechanisms are neither appropriate to natural scenes nor correct models of human perception. In one, a scene is lit from one side by one colour and from the opposite side by another. The lighting near the centre would roughly correspond to the average

illumination, yet compensating by this amount at either edge of the scene would give too much of one colour and not enough of the other. A second experiment is that even a single green light from the side would mean, after compensating for the average flux, that the objects near the light would all appear to be shades of bright green, while those far away would seem to be very dark and lacking in green. A third demonstration is based on a common problem for photographers; when sunlight spills through a window, it creates a scene that is partially sunlit and partially tungsten-lit. Unlike the first two, this demonstration does not depend on a light gradient, with some areas getting more light than others. We simply light one part of the scene with one source colour, and another with a different colour. While the average in each part might be useful for compensating in that part, the average over the whole scene is incorrect to the extent that the two colours are different. The obvious improvement is to compensate at each point by the amount and colour of light at that point. Unfortunately, this is circular, because it does not provide any way to determine how much light reaches each scene point, or from which source.

The inadequacy of traditional explanations of colour constancy will be seen more thoroughly when the literature is reviewed in the next chapter.

Clearly human visual perception is doing something quite clever. It is receiving the flux from each point in the scene, where the flux depends upon both the lighting and the reflectance, yet it removes even complex lighting patterns to determine the reflectances of the objects. This ability is known as colour constancy, since objects are seen to have a constant colour independent of the light that they

are viewed under.

It is clear that both natural and man-made environments are subject to substantial changes in illumination colour, although we rarely notice the full extent of these variations, so good is our colour constancy mechanism. Consider the differences between dawn, noon, dusk and moonlight. Both cloud cover and forest canopy alter the light colour. Further, every surface can act as an illuminant; objects near a chalk cliff get a second, yellowish, light from the cliffs, as well as the primary light directly from the sun. Any object acts as a secondary source to the objects around it, with the amount of effect depending upon its size, lightness and distance. In man-made environments, there are incandescent and fluorescent lights, secondary reflectance from coloured walls, ceilings and floors. Consider neon lighting, discos, street lighting, television pictures, and industrial and experimental illumination.

It is also clear that determining the object colour is environmentally useful, since it improves object recognition and identification of materials. It helps distinguish between food, inedibles and poisons. It separates a striped yellow tiger from the equally bright foliage. It would help identify both species and individuals.

Land [1959a,b and elsewhere] has shown experimentally that human colour constancy is very robust, perhaps beyond the needs of natural environments. In one case a two-dimensional scene composed of coloured rectangles was viewed under a flash of coloured light. The subjects correctly identified the rectangle colours despite there being insufficient time for eye movements, and despite having not previously experienced the scene or the illumination mixture. The

abstract rectangles remove the possibility of colour recognition based on object identification, such as seeing a banana and then knowing it is yellow. The scene is flat because of a decision to study only two dimensional worlds. However, it also avoids shadows and curved surfaces, which we will later see as providing strong colour constancy clues.

In another experiment[1959a,b], a range of degenerate lighting was used to determine the characteristics of the minimum illumination sufficient to produce colour constancy. It was found that a single wavelength, or spike, of any frequency was insufficient to generate a full range of colours. It gives instead a monochrome perception with a general tint of the colour of the illumination. This is why many colours are hard to distinguish under sodium street lighting, which has a narrow dominant frequency; with only one frequency there is only one dimension of reflectance information available to the eye. However, two spikes that are not too similar in wavelength do produce coarse colour identification, allowing main colours to be distinguished, but not always small colour differences. This work is important, since the optimal choice of wavelength for two sources gives colour properties adequate for some applications, and so a two-emulsion film might be made in addition to the more expensive, conventional three emulsions.

We will see later a theoretical argument that demonstrates that either three or four independent colours are required for complete colour matching in humans, depending upon the role played by the rods. Those familiar with high quality colour reproduction may note that often seven or more pigments are required to reproduce a scene well. The reason for this is that the colours of the printing

pigments are not the same as those of the human retinal pigments. If economic pigments of the same spectral response as retinal pigments were available, then only they would be required for easy and exact photographic reproduction. For experimental purposes, Land has created such pigments for use in his Retinex filters.

Now that the nature of the problem has been seen, the related technical terms may be defined. Reflectance is a property of objects, indicating what amount of the light that strikes it is re-transmitted. Flux is an amount of light. It will be used for the light available to the eye, and sometimes for the light reaching an object. Illumination refers both to the sources of light and the light itself. A point source is a comparatively small illuminant, such as the sun or a tungsten bulb, and unlike fluorescent strip lighting.

Reflectance is wavelength-specific, in that an object may reflect more red than green. This, indeed, is the reason for object colours. Non-specific reflectance, indicating the amount of white light reflected, is called lightness. Since, ultimately, "white light" is only definable in perceptual terms, and because "lightness" is commonly used with reference to the amount of light in say a red receptor channel, lightness is technically definable as a scalar amount of reflectance.

The topic of the present work would most appropriately be called reflectance constancy, but this name is not in general use, with colour constancy being more common. However, there is a large body of work dealing with the sub-problem of lightness constancy, which can be taken as colour constancy in the one dimensional reflectance world of white-grey-black. Brightness, in contrast, is a property of

flux rather than reflectance.

Colour is used herein with three basic meanings. The first is the everyday meaning of colour names such as orange and yellow. The second is as a property of reflectance or flux that indicates the amount at each wavelength. This might be expressed as an intensity curve over the visible spectrum, giving the reflectance or flux at each wavelength. However, no biological or mechanical vision system measures every wavelength, instead a few types of receptors are used, each with a different spectral sensitivity. In the present work therefore, colour is usually expressed as the set of values given by the receptors at that point, such as 12 units of blue and 18 units of green. A common system for colour notation is called RGB, because it measures the respective amounts of red, green and blue light. The third usage of colour is similar to the second, except that it excludes the black-white component. With this meaning, flux of 6 units of blue and 9 of green would have the same colour as the example above, but only half the brightness. The appropriate meaning of colour will be made clear in each context.

A pixel is a single point in an image. The word is a conflation of "picture cell". Although we see the world as continuous, the eye is actually sampling it at discrete points, where the light strikes a rod or cone in the retina. For the eye, each receptor is a pixel. The pixels of a television picture are easily seen on close inspection. Practical computer vision at the moment is about four times coarser than a television picture.

The word scene will be reserved for the three dimensional world of objects. Image will only be used for the two dimensional picture after it has been captured by the eye or camera. Edges are usually

properties of scenes, where two regions meet. However, edges also occur in scene objects where two faces meet. A junction is the meeting of three or more edges in a scene. When one object in a scene is in front of another, then the edge where they meet is said to be an occlusion. Occlusions are of interest because the regions are adjacent in the image, but separated in the scene.

Reflectance may range between being specular or matt. Specular surfaces are glossy, producing surface highlights. The most specular surfaces are like mirrors. In contrast, the reflectance of a matt surface occurs within the material rather than on its surface. Fabrics, for example are usually matt. A specular reflection does not change the colour of the illumination, but a matt reflection does. Therefore, matt reflectances, which are the most common, are of most interest in colour constancy. A Lambertian surface is a matt region which reflects in exactly the same amount in all directions.

Occasional reference will be made to CCD cameras. These are a new type of television camera, which have somewhat different properties from ordinary TV cameras. This is because they have a separate light receptor for every pixel of the image, whereas a conventional videcon has only a single receptor per colour which then scans the entire image several times per second. If the reader is unfamiliar with these cameras, the brief references to them may be omitted.

Although every attempt has been made to reduce the mathematical content of the presentation, some terms are helpful in avoiding repeated descriptions of a property. The surface normal is what one might think of as the perpendicular to the surface. On a flat, that is planar, surface, the normal remains constant for every point on that surface. If an object curves, then the normal changes gradually

between adjacent pixels, whereas at an object edge, such as between faces of a cube, the surface normal changes abruptly. Vectors are used in the sense of an ordered collection of numbers. For example, a (red green blue) vector might have values (15 8 12). Each value in a vector may be called an element or a component. Vector arithmetic is not necessary for our purposes, but in chapter three a component-wise multiplication which is inherent to colour vectors will be defined. An ordinary number, that is not a vector, is also called a scalar.

When considering colour constancy, it must be kept in mind that a colour constancy algorithm can always be fooled. Consider a photographic print. When we look at it, we perceive objects of different colours, some in shadows, some in bright light, some perhaps under trees, and so in a different colour of light. The same things may be perceived when looking at a projected photographic transparency. In both cases we perceive a rich scene of both changes of object reflectance and changes of illumination. Yet in case of the print, there were actually only changes of reflectance, and in the transparency only changes of illumination. In both of these cases colour constancy can be said to have failed. If it worked for the print then the regions perceived as in shadow would actually be seen as a separate region of different, darker colour, since the object, which is the paint, is actually darker. If reflectance constancy worked for the transparency, then only one object would ever be seen in a slide show. That object is the white screen itself, because the shapes projected onto the screen are not reflectances, but mere lighting phenomena. These may seem to be extreme examples, but they show that any perception is obtainable,

given control of the flux independently at each pixel. Indeed, it is only because of the clever heuristics that the eye employs in colour constancy that these photographs are perceived correctly, rather than literally.

Although it is not always acknowledged, colour constancy is closely related to the phenomenon of lightness constancy. The latter term came originally from processing monochrome pictures, and was in fact a single-colour "colour constancy". That is, an object retained its "colour", a single grey-scale value, independently of illumination. Here we use it with a subtle but important difference. It is the ability to tell how much of the difference in brightness between two points is due to the difference in their reflectivity and how much is due to the difference in the amount of light that reaches them. It assumes that either the objects are the same colour and under the same colour of light, or some other algorithm has compensated for the different colour of surface or illumination. Much of the work in edge detection is bound up in lightness constancy, attempting to tell whether the flux change between pixels is due to a change in incident light, so that the points are part of the same object, or whether the difference is due to a different reflectivity, meaning that they are part of different regions and separated by an edge.

The present work is an extension of the work of Edwin H. Land and his colleagues at Polaroid, and his nominal opponent, Alan L. Gilchrist. Professor Land is an experimental scientist and an inventor, and has greatly contributed to the understanding of the nature and importance of colour constancy. Unfortunately, his

writing is not as strong on theory, and he has limited himself to two
dimensional scenes in order to avoid the complexities of lighting and
surface phenomena in the real world. In the present extensions of
his algorithm the structure has largely been retained, but each
component has been replaced, and some new steps added.

Alan Gilchrist is known as the man who makes black look white.
This is because in a series of experiments [Gilchrist, 1977, 1980a]
he has created scenes where observers consistently make dramatic
errors in identifying the colour and lightness of surfaces. It will
be seen later that these errors depend upon extended coincidental
alignments within the scene, and an unusual impoverishment of the
objects. Nonetheless, studying scenes in which reflectance constancy
fails is as important as studying scenes where it succeeds. A major
success of the present algorithms is that they are as good as humans
at interpreting Gilchrist's scenes, unlike any previous algorithm.

In the results of the present study, there are two major
surprises. The first surprise is that, in colour scenes, edges do
not have nearly the importance that previous researchers, in both
monochrome and colour, have ascribed to them. In fact, a complete
colour constancy algorithm for one domain of scenes, given in chapter
four, does not require any use of edges at all. In general it turns
out that colour information gives easy and accurate region groupings,
and edge measurement is only necessary when considering lightness
constancy. Nonetheless, perceptual evidence suggests that edges are
quite important in the functioning of the human eye. This is
probably a computational convenience. The present work does not
claim to report how the eye performs reflectance constancy, but
rather to show that there are several cues to reflectance that have

not been previously discussed, and that a reflectance constancy algorithm using these cues is more versatile than any previous man-made algorithm.

The second surprise is that, within the proposed algorithms at least, scene complexity often reduces the difficulty of colour constancy, rather than increases it. For example, a complete colour constancy algorithm is given for an arbitrary scene that is lit by three or more independent colours. No version of this particular algorithm will work with less than three source colours. Similarly, curved objects in the scene generate a variety of fluxes, and these are not needless complications to be filtered out, but sources of important information for separating region and lighting colour. Similarly, the techniques developed here are tolerant of, if not enhanced by shadows, large objects, extreme object colours, and partially obscured objects that appear as two or more disjoint regions. When two or more illumination colours occur, the information obtained in one can be immediately used in the other.

Chapter two details the work of both Land and Gilchrist, analyses Land's algorithm and demonstrates its weaknesses. It is seen as consisting of several modules combined in an overall framework. The modules are especially deficient in ignoring most of the colour information; they are essentially monochrome. The overall framework can be seen as caused by the need to separate brightness steps from colour gradients. As it stands, it can not be extended to three dimensions, but if the modules are changed, a similar framework can be used to first determine the colour differences between regions, then the lightness differences, and finally the actual scene reflectances.

- 12 -

Chapter three introduces what may be called flux space theory. One concept of this theory is the dimension of the illumination, which is the number of light sources of independent colours. When a scene is viewed by a number of colour receptor channels greater than the dimension of the illumination, the theory shows how to determine the number of source colours, and to unambiguously separate the effects of each source colour. It can predict the flux that would be seen under any one of the illumination colours in the absence of the others. It also allows, for almost any scene, the determination of the ratio of surface colours between any regions of the scene. This can reduce the problem of reflectance constancy to one of lightness constancy.

An essential part of flux space theory is that it is not chained to the spatial information. When projected into flux space, pixels from a region will group themselves into a flux structure, comprising all the points with the same reflectance colour, independently of their location in the image. This provides immediate region detection, as well as unifying disjoint regions of the same colour. It also means that error reduction can be done on a region as a whole, and sometimes on the entire scene, rather than on some arbitrary small mask.

After the initial grouping of regions into structures, computation continues using the flux structures themselves. One such computation is explored in chapter four, which confines itself to the world of scenes illuminated only by distant point sources. Within this world the flux structures are limited by a function named the envelope. When two or more sources are present, this envelope allows computation of the relative angle between the sources, and their

relative lightnesses. The envelope also provides increased information about the lightness of the surface and its illumination. Finally, the envelope techniques can determine or greatly constrain the surface normal at each pixel. That is, the angle of the surface of the object at each point is known. Other researchers have shown how to use this information to determine or constrain the depth of points in the scene.

When three or more colours reach the scene the remaining ambiguities are removed, because every pixel then lies on the envelope. This is a much more powerful condition, and provides a complete colour constancy algorithm which does not need any edge techniques or external lightness constancy information.

Chapter five returns to general scenes to show how to combine flux space theory with externally determined lightness measurements to complete colour constancy. It shows how to transfer information between source colours to improve the results in each. It also shows how to combine edge and flux space information to determine whether or not an edge is due entirely to reflectance change, because this is the criterion for edges that can be used for lightness measurement. Finally, it shows that some edges are more important than others in colour constancy, and shows how these edges can be found for quick approximation algorithms.

Chapter six looks at an entirely different method of determining surface reflectance, and so reflectance constancy, using three extreme forms of secondary reflectance: ideal background illumination, mutual reflectance, and veiling luminance. The background case is a comparatively useless computational curiosity, since it assumes that every point in the scene receives the same

amount of background illumination. Nonetheless, it shows that using flux space theory, such a background would immediately give lightness constancy for every region, and thus complete reflectance constancy.

In mutual reflectance, two objects each reflect enough light to act as a significant illumination source for the other. This case is studied with particular reference to corners where two surfaces meet. A method for extracting the colour and strength of the illumination is developed, using only information local to the corner. This is enough to complete scene-wide colour constancy. It is seen that if spatial information is known about the corner, then more can be determined directly.

Perhaps the most important part of the chapter is the detection and removal of veiling luminances. These occur when every point in the image has a constant flux added to it. It occurs when looking through a shop window or through the surface of a still pond, but more significantly, it constantly occurs in human vision because of the light scattered and reflected within the eyeball. Interestingly, Gilchrist [1983b] has shown that humans can only detect and remove the veil when there is a substantial complexity in the scene, as is the case outside perception laboratories. No previous theory explained how such a veil could be processed. The success of Flux Space Theory in removing the veil was not a goal, but rather an unexpected corollary.

In chapter seven, it is seen that colour information is well suited to intelligent error correction, and interpreting mixed pixels. Colour can also allow photometric stereo to be performed on moving scenes. Photometric stereo is a method of estimating depth by sequential control of separate light sources. The flux space and

- 15 -

envelope theories suggest a number of perceptual experiments, some
intended to determine how well these algorithms model the human
visual system, and some to better understand the interaction between
lightness constancy and colour constancy.

Land and Retinex Theory

Gilchrist and False Perceptions

Other Literature

The present work is set in the context of thirty years of development of Retinex Theory by Edwin H. Land of Polaroid and his colleagues [Benton 1977, Daw 1962, Frankle 1980, Land (all), McCann (all), McKee 1977, Stiehl 1983]. Their results are summarised in Land [1977b] and Land and McCann [1971].

The name Retinex comes from the separate channels which Land postulated for colour processing. It is a conflation of the words retina and cortex, because he wished to indicate that the processing which he described might take place in either or both areas. This notion of keeping the channels separate is one of the important contributions of his theory towards developing colour constancy mechanisms.

Early in the 1950's Land was experimenting with colour separations as a prelude to extending one-minute photography, as it was then, into colour. He had three projectors in registration on the same screen. Each had in it a monochrome picture of the same scene, but one photograph had been taken through a red filter and was also similarly projected through red. The other two were respectively photographed and projected through green and blue. In the course of using this apparatus, one day the blue projector failed. Land happened to notice that the colours on the screen were substantially unaltered. Being well versed in colour mixing theory, this troubled him. If there were only shades of red and green on the screen, then

the perceived colours should not include blues, browns, yellows and purples. The evidence before him was wildly divergent from established wisdom. This chance occurrence, and the attentiveness to notice both the perception and its related paradox led to thirty years study of colour perception, and to the evolution of Retinex theory [Walls, 1960].

He next discovered or was shown the similarity between his discovery and the two-separation colour systems of the early cinema. Due to the necessarily short exposure time of each cinema frame, as well as the complexities of development, colour cinema became practical much later than still colour photography. Nonetheless, there were many imaginative attempts at colour cinema which are now almost totally forgotten. Fox and Hickey [1914] used a system where the scene was alternately photographed with and without a red filter, and then projected with the same alternation. Later, Bernardi improved the result by adding a green filter to the non-red photographic stage, while retaining white light for projection [Cornwell-Clyne, 1951]. The Cinechrome system of 1921 used a double width film with the colour separations beside each other for simultaneous projection, and there were many other variations. Evans [Land, 1959d] used red and yellow light to obtain full colour in still projection. The reason that these techniques work will become apparent as we examine Land's experiments with them.

Textbook colour mixing theory says that if red light is added to white light, the result is pink. With an image in each light path, giving different amounts of red from one, and different amounts of white from the other, the predicted result is a variety of shades of white, pink and red. Similarly, red mixed with yellow is expected to

produce only shades of red, orange and yellow. However, this was not the case for these early experimenters; their apparatus produced nearly the whole spectrum of colours as seen in the original scene.

Land wished to resolve this paradox, and to determine the capabilities of the process [1958,1959a,b,c]. His first public demonstration was May 1955 [Walls, 1960]. He provided a series of demonstrations that the process was quite robust, and that wavelength does not determine the colour perception. The latter contradicts the assumption, common since Newton's discovery of the spectrum, that the eye measured object colour by determining the strongest wavelength reaching the eye, perhaps modified by a wide variety of factors such as adaptation and contrast.

Most of Land's early experiments use two monochrome transparencies, taken identically except that one, the "short record", is taken through a green filter, and the other, the "long record", is taken through a red filter. These are projected through two projectors in registration onto a single screen. Since each is monochrome and both are projected in white light, the effect is a monochrome image of the original scene. The short record is now turned off and a red filter added to the long record. The result is the same monochrome image, but seen with a red "wash". When the short record is again illuminated, the expected result from combining the two projectors, one of white light and the other of red, is a mixture of red, pink and white. However, this is not the case; the image is seen in full colour, with each original hue correctly perceived. It is important to note that each time the image is presented, the colours appear immediately and correctly, thus eliminating explanations based on bleaching, adaptation, and other

- 19 -

time-dependent visual phenomena.

The experiment may be repeated using a dual monochromometer. This device allows a narrow wavelength band, as opposed to the broad bands used above, to be presented independently to each image. Land has shown that if the two wavelengths are not too similar, then full colour still results. Most impressively [1959c], if one wavelength is a short yellow and the other is a long yellow, then the eye still sees all the correct colours, even though it is only receiving a mixture of two shades of yellow. Neither colour mixing theory nor Newtonian spectral theories would predict the perception of reds, blues, browns, greens, and other colours, from only yellow light.

It is important to avoid the notion that this is some kind of trick. While examining the development of Retinex theory, it becomes clear that these phenomena are simply part of the human colour constancy mechanism. This mechanism extracts the correct surface colours from the scene, despite unusual illumination, such as the two wavelengths of yellow cited above. An important difference is that, with various tricks and illusions the perception differs from the reality, whereas here the colour perception is the same as the reality, which is the reflectances of the original scene.

Several experiments demonstrate the robustness of two-channel colour. Wide variations, up to a hundred fold, in the ratio of light from the projectors do not substantially alter the perceived object colours. Colour mixing theory would say that as the red projector gets brighter, the redness of the objects of the scene would increase. If the room lights are partially illuminated so that the screen and the room can be perceived simultaneously, the colours of both are seen correctly simultaneously. A second set of

monochronometers may be added, adjusted to a different pair of wavelengths, giving three simultaneous and correctly perceived "colour worlds". This is related to the ability to correctly see object colours in a room which is partially sunlit and partially artificially lit.

It is possible to double the contrast of the image by putting two copies of a negative together in the same projector. This greatly alters the transmission ratios between regions in that image. Doing this to one or both projectors does not alter the perceptions. Although Land does not note it, with hindsight this can be seen as evidence that colour constancy processing in humans is done on the log of the flux, rather than on the flux values themselves.

Land suggests that the two colours of the monochromometer provide two "arbitrary primaries", in that all the perceived colours are produced by an amount of one added to an amount of the other. Evidence for this occurs when the two transparencies are exchanged. When the long record has the shorter of the two wavelengths, then full colour is still perceived, but in reversal. The perception is that of a colour negative; red has become green, yellow has become blue, and so on. This also demonstrates that the perceived colours are independent of any expectations, indeed stronger. Independence is also shown by using photographs of abstract coloured papers, or objects which do not have a predictable colour.

Land's explanation of how correct colour perceptions are obtainable in these experiments is that the two photographic transparencies store different information about the scene colours. When the wavelengths that carry them are perceived, it is not by choosing the stronger of the two, but rather by using them separately

as co-ordinates in a two-dimensional Cartesian space of colours. The location of each reflectance colour in this space is computable from the images themselves, as the percentages of transmission, and this has been done [Land, 1959a page 125, and elsewhere]. A virtue of this model is that changing the amount of light from one projector causes the dimensions of that axis to be rescaled accordingly, without altering the topology of the colours in the space. Similarly, Land shows that other alterations to the flux which preserve the topology, such as change in contrast, do not reduce the reflectance information available.

The essence of colour constancy in these demonstrations is that the original photographs have respectively captured the proportion of long and of short wave reflectance of each object. These are then re-transmitted on any two carriers which preserve the long and short wave relationship, and are far enough apart for the eye to separate. After separation, they act as indexes to the colour space, thus specifying the object colour. Just as the wavelength of a radio station is immaterial when listening to its programme, so the wavelength of the two light carriers is immaterial to the colour perception. This analogy is especially good with respect to stereo radio, where two adjacent wavelengths carry different but related information which is perceived by comparing the two channels.

Land does not explicitly address the issue of how the eye separates the two wavelength carriers. This is not a contentious issue, but it should be explained for the reader. Essential is the fact that the three cone systems are broad band, and have different peaks. Any colour of light, at reasonable intensity, will stimulate all three systems, but in differing amounts. Say that wavelength A

stimulates the red cones more than it does the green cones, and that wavelength B stimulates the greens more than the red. An increase in A will result in an increase in both red+green and red-green, while and increase in B will result in a decrease in red-green but still an increase in red+green. A decrease in either will invert the sign of red-green, and decrease red+green. In short, comparing the cone systems separates the two carriers in much the same way that comparing the two signals specifies object colours.

In addition to showing the remarkable versatility of colour constancy, some of Land's demonstrations define conditions which do not produce object colour perception. Most of these are cases where the flux points that occur on the screen all fall onto a single line in the colour space. One example is when the same transparency is used in both projectors, or when one is the negative of the other. This observation is central to further development of colour constancy understanding; if the colour information in the image is one dimensional, then it is monochrome, but if it spans two or more dimensions, then it allows perception of object colours.

However, some demonstrations of the failure of creating colour are not caused by single dimensioned information. One of these is when the two projected images are out of registration. Even a small displacement on the screen of one picture relative to the other causes object colour to be lost, and then only the limited colours predicted by mixing theory are seen. Land says that this is because the perception of the objects is lost. However, this explanation is not acceptable. Firstly, slight change of registration does not destroy the object perception, but causes only a blurring of the edges; the objects in the image are still wholly identifiable long

after their colour is gone. Secondly, this explanation is at odds with the evidence that object recognition is not necessary for the colour constancy, which works with even abstract shapes.

A second important experiment where full colour does not occur is when the transparencies are replaced by neutral density step wedges. These are filters consisting of parallel bands of grey, increasing from transparent on one side to black on the other. With one projector transmitting through a vertical wedge and the other through a horizontal wedge, an array of square patches appears, each having different co-ordinates in colour space. These points are distributed throughout the space.

Similarly, full colour is absent when continuous, non-step, neutral density wedges are used. These have a continuous gradient from clear to opaque in one dimension while being constant in the other. Two crossed wedges produce every flux point in the two-dimensional colour space. Both examples produce two-dimensional flux information, yet neither produces any object colour effect. Land's explanation for this is that the patterns are "too regular"; the eye needs a certain amount of "randomness" to work. That is pure hand-waving, and he quickly goes on to another topic, but we shall now consider it more fully.

In later writing [Land and McCann, 1971] he demonstrates the importance of edges in lightness measurements. From that generally accepted observation it is clear that continuous wedges would not show colour constancy because the image would be perceived as depicting no more than one object. As an aside, the flux space theory developed in the next chapter would indicate that the scene would be interpreted as unreal because of the impossible changes in

the two apparent illumination gradients on the one object.

Nonetheless, edges are present in the step wedge case, and yet object colours are not perceived. Nothing in Retinex theory up to the present time accounts for this failure. With hindsight, the weakness of the theory which prompted the present work is also responsible for the failure to explain this particular experiment. An important characteristic of the Retinex algorithm is that lightness is computed separately in each of the colours, with the results being compared at the end of the computation and serving as indexes into colour space. This is a major improvement from previous algorithms, which computed a single value, "hue", from red, green and blue at the start of the process, thus reducing three dimensions of information to one dimension and losing the other two-thirds of the data. However, Land has gone too far in separating the three colour planes. Interaction between the planes which does not lose information can greatly enhance the computation of colour constancy, as is shown in subsequent chapters. One small case of this relates to the step wedge experiment. Land has always tacitly assumed that when an edge is detected in one colour plane, then it will be detected at exactly the same place in all the colour planes. Researchers who have worked in practical computational vision would probably not have made this mistake, since practical edge detection is difficult and unreliable. As a simple example, the colour images from most contemporary laboratory devices are often out of registration, and the correspondence of edges between colours can be used to correct the alignment. This is a non-trivial task since the error changes across the image, but methods similar to stereopsis computations can be applied.

One of the early steps in the present research was to consider ways of detecting edges using the three colours simultaneously, assuming no registration problems. Clearly this edge detection is more appropriate than Land's monospectral version, since real object edges affect all the reflected light, not just the light in one receptor colour channel. Realising that edges are inherently multispectral, a better explanation for both the step-wedge and unregistered image failures would be that the edges occurred in only one channel, and thus they would not be perceived as object edges for the colour system.

This should be developed further, since a true reflectance edge can happen to be constant in one colour, although it is very unlikely that it will be constant in more than one. The solution must await the presentation of Flux Space theory.

In fact, researchers replicating Land's work have inadvertently shown that his non-random hypothesis does not hold, because they easily achieved full colour in a regular array of square patches. Pearson, Rubinstein and Spivack [1969] did this with a ten by ten square quilt of areas of constant reflectance, which was spatially quite similar to the step wedge image. The experiment was designed to show, using Land's methods, that the earlier Helson-Judd formulas would "predict" Land's results. Their success rate varied from sixty-one to ninety percent, and they regarded this as a good prediction rate.

Karp [1959] replicated Land's key experiments, and also showed that the perception of "full colour" was preserved in photographs of the screen. He noted that a pinhole view of the screen reduced the perception to the colour predicted solely by colour mixing,

confirming the importance of context rather than flux. He then used a scene of a five by five array of square colour patches on either a black or white background. The patches are nowhere adjacent because of the separation of the background. With the black separation, only mixture colours were seen, as in Land's crossed step wedges. However, full colour occurred with white separation. He suggests that this might relate to the apparent whites of the highlights seen on glossy objects. This is not supported by his further finding that even a dark grey field preserves full colour. A more likely explanation is that the black field isolates each region into a spot colour, so that no relationship between the regions is seen. Computationally, this is because the ratio with black is undefined, being essentially division by zero. Thus the ratio can not be propagated across the gap to compare the separated regions.

Karp's paper goes on to make an important note on Land's short wave reversal, which is a small group of wavelength pairs which generate an unexpected negative image when used for two-colour projections. The image is correct when the two wavelengths are interchanged, with the longer one illuminating the short record. Karp notes that these pairs correspond to unusual angles in the CIE chromaticity diagram, which would be associated with a change of role for the two colours. His paper also anticipates the Retinex concept of the virtual white, by noting that the warmest image point is the brightest in the long record, the coolest is maximum in the short, and white is the maximum in both records. Unfortunately, alongside these virtues, his paper matches contemporary feeling in attributing the phenomena of two-primary colour to simultaneous contrast, and dismissing their importance.

Daw [1962], on the other hand, was firmly in Land's camp, and was briefly a co-worker in his lab. His rectangular array, like the ones used by Land at that time [Land 1962a], had a mid-grey background, with no apparent loss of colour perception. This particular study by Daw was intended to show that successive contrast could not explain colour constancy, but the results also indicate the importance of edges in human vision. Subjects were asked to fixate for twelve seconds on a particular point on the screen during two-colour projection of a natural scene. Then the red projector was turned off, leaving a single black and white projection on the screen. Wherever the subjects looked they correctly perceived only the monochrome, except when they returned to the fixation point, where they immediately saw the full colours as during dual projection. The eye could depart and return any number of times. Full colour always recurred at the fixation point until the fading of the afterimage, which lasts about as long as the fixation period. A version of the successive contrast hypothesis is that the eye compares the flux at a point with the afterimage from previous points to somehow eliminate the effects of the illumination. It is not clear how the afterimages of green fields and white clouds are both meant to have the same effects.

Daw then repeated the experiment using an abstract scene consisting of a rectangular lattice of colour patches on a middle grey background. The fixation point was the upper left corner of a particular patch. The same results occurred, with a significant addition. Full colour would occur whenever the eye stopped at any upper left corner. Furthermore, although the colours were correct when looking at the fixation corner, they were different for each

corner chosen. There must be two phenomena involved in these results. Firstly, it is clear that edges are important for eliciting the afterimage. Traditionally, afterimages are looked for on a blank field after fixation. Subjects in this experiment said that the new afterimages were substantially stronger that those on blank walls. When the edges coincided, the perception was said to be real rather than ephemeral. The images would occur after only one second of fixation, although they would correspondingly last for only a second. With further experiments, Daw suggests that when edges are present, they tend to suppress afterimages as an aid to perception, except when the edges accidentally coincide with those of the previous image. The observed enhancement of afterimages is probably an artifact of that method rather than an important tool. Edges rarely coincide as the eye roams natural scenes, so the new edges can suppress afterimages.

The second phenomena composite in the data is the production of colours from a combination of a monochrome image and an afterimage. Recall that colour occurs when there are two or more channels of flux available for comparison. The data implies that the afterimage provided one dimension of information, while the input sensations provided the other. It is not sufficient to say that the edges of the monochrome image permitted full colour to return, because this would say that the monochrome image was ignored other than its edges, and also contradict the data of the second experiment. Daw's report is not sufficient to determine whether the afterimage simply replaced the missing red long record, or whether some other processes were occurring. In either case, the exact mechanism remains unclear. Any explanation must also take account of Zeki's finding [1983a] that the

corner chosen. There must be two phenomena involved in these results. Firstly, it is clear that edges are important for eliciting the afterimage. Traditionally, afterimages are looked for on a blank field after fixation. Subjects in this experiment said that the new afterimages were substantially stronger that those on blank walls. When the edges coincided, the perception was said to be real rather than ephemeral. The images would occur after only one second of fixation, although they would correspondingly last for only a second. With further experiments, Daw suggests that when edges are present, they tend to suppress afterimages as an aid to perception, except when the edges accidentally coincide with those of the previous image. The observed enhancement of afterimages is probably an artifact of that method rather than an important tool. Edges rarely coincide as the eye roams natural scenes, so the new edges can suppress afterimages.

The second phenomena composite in the data is the production of colours from a combination of a monochrome image and an afterimage. Recall that colour occurs when there are two or more channels of flux available for comparison. The data implies that the afterimage provided one dimension of information, while the input sensations provided the other. It is not sufficient to say that the edges of the monochrome image permitted full colour to return, because this would say that the monochrome image was ignored other than its edges, and also contradict the data of the second experiment. Daw's report is not sufficient to determine whether the afterimage simply replaced the missing red long record, or whether some other processes were occurring. In either case, the exact mechanism remains unclear. Any explanation must also take account of Zeki's finding [1983a] that the

- 29 -

colour of afterimages is dependent upon the perception, rather than the the flux of the fixated scene.

When Land's articles first appeared, they caused quite some controversy. The popular press, for example Bello [1959], immediately made predictions of two-colour photography, television, printing, and so on, and stated catagorically that all previous researchers were fools. Although Land himself made efforts to prevent such absurdities, these popular reports did much to entrench other vision researchers against him. Wilson and Brocklebank [1960, 1961] were wholly condemning, and reviewed only Land's detractors. Wright [1959], in a summary of areas of vision research, only mentions Land, but complains that "we need some way to measure Land's colour appearance". This shows that he has entirely missed Land's point. Contemporary vision research was essentially interested in studying the failure of the eye; a stimulus would be presented that would seem like something else, and so it was important to measure the false perception. Land insisted that the important behaviour of the eye is not in its failures, but in its broad range of successes. The essential point of the two-colour projections is that the correct object colours are perceived. Wright, however, insisted on regarding them as simply another visual trick, in which red and green are added to create the "false" perception of blue, and so on.

Judd [1960] made a well thought out appraisal of Land. First, he showed that "classic theory", as Land had termed it, is not limited to colour mixing. Indeed, Helson's formulation, with Judd's own modifications, predict the colour perception from mixtures of colours in different proportions, including results unexpected from colour

- 30 -

theory.   Nonetheless, he accepts the value of Land's work on several
counts.   Land replied [1960], accepting that   the   two-colour
projection techniques are not new.   In fact, he had cited early
cinematographic references that even precede Helson.   However, he
indicated that all of his previous points stood, including colour
perception coming from channel comparisons rather than mixtures, and
the value of complex scenes in vision research.

The literature of the time has similarities with the Ptolemian
theory of the solar system and its preponderance of epicycles. In
that theory, the original hypothesis was that the   planets   travelled
in circles.   As data collection became better, the hypothesis was
modified to include tiny circles moving within larger circles within
still larger circles. There was great resistance to the notion that
the orbits might be ellipses.

In the case of vision, the initial stage was the   assumption that
the colour of an object was the colour of the flux, as with a
photometer. Epicycles on top of this vary, depending upon the vision
paradigm that one was trained in. They include lateral inhibition,
followed by another cycle of disinhibition, or the shift from flux to
flux ratios at edges. The latter, in fact, has only recently been
challenged [Gilchrist and Jacobsen, 1983a] in a way that shows that
it is grossly unable to explain reflectance constancy. The common
response to Land's demonstration that full colour can be obtained by
comparisons of only two yellow frequencies was to propose another
adjustment to the old formulas, rather than to consider a theory of
the comparison of channels.

Marr [1982] has pointed out an essential distinction between two
uses of the word "prediction" in this context. In the Helson-Judd

- 31 -

formulation, prediction means that by plotting the fluxes in their colour diagram, the approximate value of the perception can be estimated. No one suggests that this is how the human system operates; no physiologist has yet found a colour circle inside a brain. The Land system, in contrast, suggests a mechanism, and that mechanism matches the "ecology" of the task, which is the problem of determining surface reflectance in a world of greatly varying fluxes. In fact, Land's model also correctly predicts when colour constancy will occur, a prediction about the perception process, rather than simply perceptual data. Essentially, the earlier formulation succeeds because it has collected so much data that new conditions can be predicted by finding similar conditions in the data table and interpolating between values. The regularity in the data is an artifact of the regular structure of the perceptual system, and does not consist of a "theory" in itself.

These objections are not meant to denigrate the value of earlier research. Much has been learned about the actual workings of the eye. A new theory does not discard the previous data. To show that lateral inhibition does not explain constancy is not the same as claiming that inhibition does not occur, or even that it is not a contributing mechanism in achieving the constancy. The point of controversy is that earlier theories, adequate to explaining phenomena in limited laboratory environments, have not proved able to explain perceptions in complex scenes. Essentially, researchers have been asking two different sets of questions. Land's questions were about constancy in real scenes, the ability to differentiate between reflectance and illumination. Others had instead concentrated on the vision mechanism itself, systematically studying its individual parts

and stages. One valid technique for learning about a hidden mechanism is to measure its behaviour in difficult, marginal or degenerate conditions. However, information gained in this way can not usually be generalised to fully describe the central functions of that mechanism. The difference between the two approaches is almost analogous to the difference between studying how to drive a vehicle and studying how to take it apart.

Although these distinctions of methodology were not apparent at the time, some commentators were more receptive. Karp [1960] implicitly suggests that two-colour projection simulates a certain colour blindness in some way. This observation was based on the poor blue-green definition, which is thought to be similar to tritanopia. The connection is more than superficial, since several colour blindnesses seem to produce either a one or a two dimensional perceptual space, rather than the standard three dimensions.

Perhaps the fairest evaluation came from Walls [1960], who had joined Land for some time before the embarrassing popularisation of the research. Walls tells the otherwise unpublished story of Land's earliest studies. For instance, he asked that Land align a black grid to the crossed step wedges, and this caused some colour to be seen. Although still critical of Land's theory, Walls is alone in his day in addressing the issues rather than the data, and perhaps this is why most of his predictions have come true.

In their review of contemporary research in colour vision, Hurvich and Jameson [1960] devoted only a small section to colour contrast, which they took to include constancy. The sort of research which they were interested in included the perceptual results of photochemical breakdown under unnaturally blinding brightness, and

the effects of wearing spectacles with differently coloured lenses for many days. They give Land a short mention, declaring unequivocally that the results are "neither more nor less than stunning illustrations of simultaneous induction or contrast mechanisms". They go on to equate the phenomena with coloured shadows. These shadows occur, for example near sunset when a blue shadow results from the red-biased sunlight being taken for white, and so the white-ish background light that fills the shadow is perceived as proportionately un-red. Although these shadows are distantly related to two-colour projection, their behaviour is confined to the colour mixing laws, the edges are always limited to changes in a single primary, unlike reflectance edges, and the shadows are once again an example of perceptual error rather than robustness.

Later, Jameson and Hurvich [1961, 1964] reported results which seemed to show that the eye was surprisingly poor at lightness constancy on a scene of five reflectance papers of shades of grey. However, their results can also be interpreted to show that lightness constancy is the primary phenomenon, with a secondary effect of a mis-estimate of the overall brightness, with a minor third-order effect that the blackness of the darkest area increases with luminance. In any event, this one anomalous experiment has never been successfully replicated despite attempts by at least three different groups [Gilchrist 1983a].

Two decades later, they have changed their mind about reflectance constancy and begun new research. Leo Hurvich has been seen to give a conference paper wearing a very loud jacket and tie that become quite sedate under appropriate illumination. The description is

- 34 -

reminiscent of the earlier theatrics of Land and many other vision researchers. Now their complaints about Land [Jameson and Hurvich, 1983] are that he has not gone beyond Mondrians, that he looses the illumination information, and that he does not account for the errors of human constancy. The first is fair enough, except that Land's co-worker McCann [1983a] reported results on outdoor scenes in 1982. The second complaint is fundamentally correct, has also been made by Gilchrist [1979a] and others, and was a starting point for the present study. We will be interested to see any method they propose for better utilising the illumination gradients. The final complaint is entirely undeserved because McCann, McKee and Taylor [1976a] measured the human bias to constancy at different brightnesses and then built a computer program that reproduced the human perceptions including this bias. They have also corrected for the bias caused by light scattering within the eye [McCann, Stiehl and Savoy, 1979; Stiehl, McCann and Savoy, 1983].

An interesting side issue raised by Land [1965], is that a polychrome visual system can conceivably be less favourable for evolutionary adaptation than a monochrome system. He recalls that in the early days of monochrome photography, the orthochromatic films were regarded as far superior to the panchromatic, because the former gave much better contrast and so object definition. In orthochromatic, reds are dark and greens are light, while in panchromatic, they are nearly the same shade of grey. By analogy, if a creature had a red-sensitive monochrome system, and then evolved a second, green-sensitive system, and directly combined the results of the two it would actually loose much of the ability to distinguish red from green. From this he demonstrates that the common notion of

- 35 -

combining the output of the three types of cones into a single colour value is not reasonable, and that it is much more useful to process colour as a set of separate values, and so yield a result in a two or more dimensioned space of possible perceptions.

He leaves unaddressed the issue of how the evolution of a creature could make the leap from monochrome perception to colour constancy, if the intermediate step is less adaptive. We may suggest that monochrome perception has at least two developmental stages. The first is flux vision, which lacks constancy. A frog will jump when a large dark object enters its visual field, since the object may be a predator, while it will strike at a small dark object, which is likely to be an insect. This sort of vision is quite useful to the frog, even without lightness constancy. Philosophically, lightness constancy compares objects within a scene, whereas flux vision compares objects between time frames. It is only at a second stage of evolution that lightness constancy need be developed.

These same two stages can be drawn within polychrome perception. It is easy to imagine several types of multi-dimensional flux vision, each lacking in colour constancy. For example, one channel might be sensitive to the colour of a favourite food, while another distinguishes the colour of possible mates. They need not have any obvious interconnection, although simultaneous signals from both channels might have a third environmental meaning. The most obvious evolutionary issue is whether flux vision evolved into lightness constancy, which then developed additional receptors producing colour constancy, or whether flux vision became multi-dimensional before developing constancies. This could be resolved if zoological perceptionists could find species that exhibited either of the

intermediate stages, although it is also possible that both evolutionary paths have occurred. Further, it is likely that several different stages in the development of lightness and colour constancy can be differentiated, and any such results would have implications for the development of other colour constancy research.

These thoughts about ontology suggest a subtle but important distinction between two theoretical types of colour constancy. If two independent channels, each with lightness constancy, provide an index into a two dimensional space, then this is actually two-dimensional lightness constancy, and as such only approximates colour constancy. True colour constancy occurs when additional computation is done combining the different receptor systems to better define the effects of the illumination. Multi-dimensional lightness constancy is unable, for example, to correctly process simultaneous illumination by different coloured lights. To a large extent the differences between Land's constancy and the mechanisms presented in the present work parallel this distinction; Land's methods make comparatively little use of information compared between the channels.

The most important publication on Retinex theory was [Land and McCann, 1971]. Previous work had concentrated on two-primary colour and demonstrating the problem of colour constancy and the inadequacy of previous solutions. In 1971 they presented their Retinex colour constancy algorithm, which incorporated two entirely novel features, lightness chaining and the colour cube. Subsequent writing on Retinex theory has only refined details of the algorithm and presented it to a wider audience. We will now examine it.

The Retinex algorithm is confined to the world of Mondrian scenes.

These are two-dimensional scenes made up of areas of constant reflectance, and having no abrupt changes of illumination. However, the illumination may have a large gradual gradient across the scene, as for example when the light comes from one side. These scenes are called Mondrians because of their similarity to some of the paintings of the Dutch artist Peiter Mondrian. The algorithm is restricted to this domain because in Mondrians, changes of illumination and changes of reflectance can easily be separated; the former are always gradual and the latter are always abrupt. In addition, the two never change simultaneously, in the sense that there is never more than a slight change of illumination across the edge between two regions. This is not generally true in real world scenes; if one object is in front of another so that the two are adjacent in the image, then either object can be in less light than the other, or even in shadow.

Each stage of the algorithm is done entirely separately in each colour component. That is, the redness of each region is found, and the blueness, and the greenness, each entirely independently of the others. The method is equally applicable to finding the lightness of objects in a monochrome image of a Mondrian world.

The first step of the Retinex algorithm is to separate the changes of illumination from the changes of reflectance. This is done by thresholding the differences of adjacent pixels. If there is a region boundary between the two points, then the difference will be larger than if the difference is only due to a slight change in illumination. Land thus selects a threshold of the right size to distinguish these two cases, and in so doing, he has isolated all of the reflectance edges. The ratio across an edge is the reflectance ratio between the two regions. So if the ratio is 2:1, then the

first is twice as bright as the second. However, this does not give the actual reflectance of either region, since the two could equally well be white and light grey, or dark grey and black.

The next step of the algorithm is to chain these ratios in order to find the relative reflectance between distant regions in the scene. This is done by taking the sequential product of the edge ratios. So, continuing with the example above, if a third region has a ratio of 1:5 with the second, then $2/1 * 1/5$ is $2/5$, or a ratio of 2:5 between the first and third regions. It is important to understand the difference between this sequential product and directly taking the ratio of the fluxes in the two regions. When two regions are far apart in the scene, the gradually changing level of illumination may be quite different between them. Their flux ratios combine both the ratio of illumination and ratio of reflectance, and so are not a good measure of reflectance. The thresholding stage is able to isolate reflectance, but it only leaves relative edge ratios, rather than absolute region intensities. The method of sequential product allows these local reflectance ratios to be propagated across the scene, so that the reflectance ratio of any two regions may be found. This gives complete information about the relative reflectances in the scene, but as yet the overall or absolute level is unknown.

The absolute reflectance is found by the final stage, the construction of the colour cube. The essential insight is to notice that in any non-degenerate scene there is always some light-coloured object. A whitish object reflects more of every colour than does any other non-fluorescent, non-spectral object. Once all of the reflectance ratios are known, it is easy to find the region with the

- 39 -

highest reflectance. This is then taken to be the value of white, and all the other reflectances are scaled accordingly. This assumption would be substantially in error only for a scene contrived of universally dark objects. A single light-coloured object would make it correct.
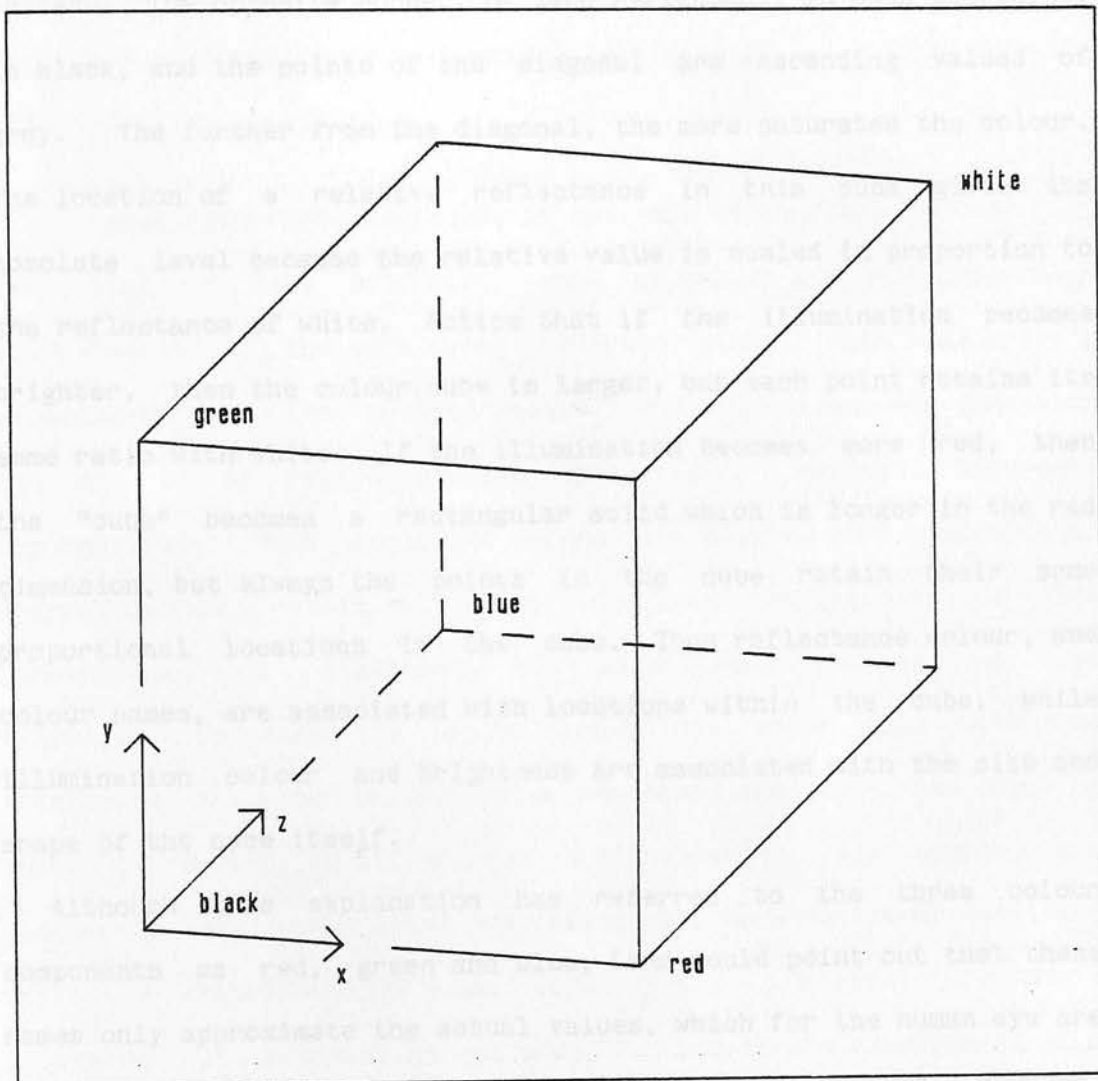
2.1 A Colour Cube

The name "colour cube" given to this scaling technique comes from the following. If each of the colour components, red, green and blue, are matched with the one of the axes of three-dimensional space, a colour cube is created as shown in diagram 2.1. Any colour triple specifies a point

- 40 -

in that space. If the smallest cube which encloses these points is found, then it passes through the most reflective surface in each colour. Conversely, finding the most reflective surface in each colour computes the dimensions of this cube. The upper corner of the cube is the computed value of white, which may be called "virtual white". The opposite corner, of zero reflectance in each component, is black, and the points of the diagonal are ascending values of grey. The further from the diagonal, the more saturated the colour. The location of a relative reflectance in this cube gives its absolute level because the relative value is scaled in proportion to the reflectance of white. Notice that if the illumination becomes brighter, then the colour cube is larger, but each point retains its same ratio with white. If the illumination becomes more red, then the "cube" becomes a rectangular solid which is longer in the red dimension, but always the points in the cube retain their same proportional locations in the cube. Thus reflectance colour, and colour names, are associated with locations within the cube, while illumination colour and brightness are associated with the size and shape of the cube itself.

Although this explanation has referred to the three colour components as red, green and blue, Land would point out that these names only approximate the actual values, which for the human eye are the output of the long, medium and short wave cone systems. However, the essence is that the cube is scaled by the receptors available to the mechanism. The cube created by a computer vision system is scaled by its own receptor sensitivities. Nor is there any magic in the number three. The same process of computing virtual white can be carried out with two or four or more colour components.

The computation of lightness was elaborated by Horn [1973]. This paper provided a more rigorous mathematical framework for the Retinex lightness algorithm. The constraints involved in choosing an optimal threshold value are made explicit. It is shown how Land's ad hoc method of chaining reflectances by a random walk can be replaced by a two-dimensional operator such as the Laplacian. Horn also shows that lightness computation is most naturally viewed as a differentiation followed by an integration.

The most important complement to the work of Land and McCann has been that of Alan Gilchrist, much of which has been undertaken concurrently with the work being reported here. Both Gilchrist's experiments and the present research studied reflectance constancy and did so by departing from traditional retinal adjacency explanations. Both projects looked at complex three-dimensional scenes, and as a result, at classifying edges. However, his work has concentrated on lightness or monochrome constancy and edge classification caused by depth perception, while this work was interested in colour or polychrome constancy and classified edges by correctly identifying differences in illumination. Furthermore, his background of experimental psychology led to a series of novel and definitive experiments in human perception. The present author's background of mathematics and computing led to the development of a computational mechanism that can separate certain features in an image into their original reflectance and illumination components in a way that had not previously been explained or even predicted.

Gilchrist's experiments serve to demonstrate that humans are able to use at least some of the information available from the Flux Space, while the mechanisms presented in the present work help to

answer the questions posed by his results.

His work was originally prompted by an observation of Irvin Rock that it is quite common for an outside corner of a building to have one side in sunlight and the other in shade, yet both surfaces are perceived to have the same or nearly the same lightness. A keyhole view of the same corner, given that the walls are plain, produces the quite different perception of a bright and a dim surface that are coplanar.

He replicated this phenomena under Laboratory conditions with several variations, and was able to demonstrate [1975a,b] that constancy is very poor if there is only one reflectance per wall, even if the spatial arrangement of the corner is correctly perceived. However, with two regions of different reflectance per wall, constancy returns. Apparently the eye needs a second surface for comparison in order to make good lightness judgements.

An important thread through this early work was to determine whether spatial perception influences perceived reflection. This issue started with Hochberg and Beck [1954], who placed a flat trapezoid upright on a floor so that it appeared to be a square lying on the floor when observed through a pinhole. The scene was illuminated from above so that the target received comparatively little light. When it was perceived as lying flat, the eye assumed that it was in full light, but when clues were added which indicated its actual position, the perceived lightness increased to compensate for the apparently lower illumination. However, in their experiments the effect was comparatively small, and others followed with a host of related experiments giving small or negative results, depending upon the disposition of the experimenters.

Gilchrist's first result explains the small size of Hochberg and Beck's lightness change. Since the target was alone in its perceptual plane when correctly perceived as upright, there was nothing to compare it with, and so it had poor constancy and tended to be incorrectly compared with its retinal surround.

Gilchrist went on to demonstrate that the change in perceived reflectance can greatly vary according to spatial perception, even with little or no change in retinal image. The coplanar experiment [Gilchrist, 1975a,1977a,1980a] presents a keyhole view of a dim room with an archway into a brightly lit room beyond. Some reflectance samples partially cross the arch from one wall, and are dimly lit. Beyond them, on the far bright wall are some other reflectance samples, so that the two sets of samples overlap in the image. A small change of an interposition cue makes one of these samples, the target, appear among either the near or the far group. When it is correctly seen as near, and thus as dimly lit, then it is correctly perceived as white. When it appears to be on the far, bright, wall it is perceived as almost black.

The corner tabs experiment [Gilchrist, 1975a,1977a,1980a] produces a similar result in an entirely different way. The subject sees a keyhole view down at forty-five degrees onto the edge where a horizontal and a vertical surface meet. The monocular view is such that only the two surfaces are seen, and they are perceived as co-planar. When a binocular view is provided, the correct spatial perception is achieved. The illumination is from above, so that the horizontal surface receives about thirty times as much light as the vertical, but no illumination gradients are noticeable. As his previous results required, two different reflectance regions occur in

each colour plane. In some cases these are small targets adjoining the central junction [Gilchrist, 1980a], and sometimes they are tabs which extend out across the view of the other plane. The latter case means that when the tabs are viewed monocularly, they are perceived as lying in the opposite plane. The horizontal tab is black and the horizontal surface is white, while the vertical tab is white and the vertical surface is black. The reflectance of the tabs is correctly perceived binocularly when the tabs lie in their correct plane, but the perception is reversed with the white surface appearing black and the black surface appearing white when a monocular view makes the scene incorrectly appear flat. In fact, the eye is receiving the same amount of flux from both tabs.

An intriguing set of experiments may be called "Gilchrist's rooms" [Gilchrist, 1978, 1984]. In these he used a pinhole view into a miniature room filled with typical objects such as tables, chairs and pictures in frames. However every surface in the room was painted the same matt colour, white for one room, black for another. Thus there was only a single reflectance within the field of view. Each room was lit by a single hidden bulb. Even though there was no second reflectance for comparison, subjects consistently identified the reflectance of the room. They could do this even when the illumination intensity of the black room was increased to the point where every surface in the black room gave off more light than the corresponding surface in the white room. Similarly, subjects correctly distinguished the white room under blue lighting from a blue room under white lighting. This is a very dramatic demonstration that perception and flux intensity are nearly unrelated. It is also a challenge to Gilchrist's own coplanar ratio

hypothesis, but he does not discuss that point. He does provide a partial clue to how the eye may do this by noting that the histogram of the black room is more varied than that of the white room.

In chapter six we shall develop two methods for solving Gilchrist's rooms. Both use the colour and intensity of the secondary illumination. The first is the depth of the shadows and the level of background illumination. The second is the amount of mutual reflectance between two surfaces that meet in an inside corner.

The newest of Gilchrist's discoveries, and perhaps the most provocative, is the nature of veiling luminance. This occurs when a constant amount of light is added to every point in the image. It occurs when one looks though a shop window, when evenly reflected skylight is seen at the same time as the objects beyond the glass. A veiling luminance may also occur when looking into a still pond, but notice that it is not the same as specular glare, which is local rather than uniform. The most significant case of veiling luminance occurs in the eye itself. In addition to the light focused on the retina, stray light bounces around inside the eye, and is known as intraocular scattering. This light strikes each photo-receptor in about equal measure, but that level is controlled only by the amount and colour of the average flux of the image.

Gilchrist and Jacobsen [1983b] studied the human ability to detect and eliminate the veil. The apparatus was a large sighting box through which the subject could see the target scene. Diagonally in the box was a piece of glass which reflected light from a diffuser and so caused a veil. In the first experiment there were three conditions: no veil, a complete veil of the same brightness as the darkest target

- 46 -

region, and the same veil with a missing, non-veil, border around it to make the presence of the border immediately perceptible. The targets were surfaces in a real outdoor scene. Subjects made Munsell matches under each condition, which were presented in different orders to different subjects. The results were that the veil, with or without borders, did not significantly alter the object perceptions. That is to say that colour constancy held through the veil.

This is very remarkable. If a photometer is used to compare regions with and without the veil, it is immediately found that every edge ratio is changed by the veil. Recall that every modern reflectance constancy theory, including Retinex theory and Gilchrist's own theory, relies on edge ratios. Gilchrist goes on to show that the reflectances computed using these ratios are wildly in error. Obviously even the best theories are not yet as good as the eye in performing reflectance constancy.

Gilchrist's second experiment eliminated the possibility that familiarity with the scene was used by the subjects. The target scene was a still life of abstract objects such as cubes and cones of varying colour. Again observer matches were consistently the same as the reflectances of the objects and different from the results predicted by existing theories.

The final experiment was to view a Mondrian scene. He tried both a monochrome Mondrian, composed of only blacks, whites and greys, and also a Mondrian with some colour regions. In both cases, unlike the previous experiments, observer matches with the veil were entirely incorrect, and matched the results that would by computed by a Retinex algorithm! Furthermore, subjects reported being unaware of

- 47 -

the presence of the veil. This has several repercussions. It indicates that Mondrian worlds are in some way qualitatively impoverished with respect to real scenes. It also shows that some aspect of the complex scenes which are universally abhorrent to perception researchers is essential for complete, correct human perception.

Gilchrist reports these results with surprise, and does not attempt an explanation. It will be seen at the end of chapter six that correct processing of veils is a natural consequence of the Flux Space algorithms introduced in the present work. At that point it will become clear what information is available that allows the eye to detect and eliminate the veil without significant additional processing. With that understanding it is easy to predict which scenes permit the lifting of the veil and which do not.

There have been a number of other researchers who have utilised the Retinex paradigm in their work. One of the most important of these is Semir Zeki, who has studied the neurophysiology of the visual cortex for many years. Recently [Zeki, 1980], he has used Retinex images to test whether given brain cells exhibit colour constancy. He has found [Zeki, 1983c,d] in the cerebral cortex of the monkey both cells that respond to the local flux colour and also cells that respond to the surface colour. These two types are respectively the input and output values of reflectance constancy. The stimulus image is a Mondrian illuminated with coloured light so that the flux and the perception are substantially different. From this finding, it seems likely that most of the reflectance constancy processing is done in the cortex, since the raw flux is present alongside the resulting perception. Further study will improve

- 48 -

understanding of the neural basis for reflectance constancy.

However, the importance of this work goes beyond the results themselves. His use of the Retinex paradigm is giving it respectability among physiologists. He has even replicated Land's demonstrations before the British Physiological Society [Zeki 1983b]. This should increase the rate of advancement in constancy studies as more researchers see the need to distinguish between scene flux and surface perception. Only a handful of other neurophysiologists, notably Daw, Frisby and Marr, have worked directly on reflectance constancy, and these have always approached the problem using techniques of other disciplines.

Another writer on Retinex theory is Bergstrom [1977]. He attempts to explain colour constancy by analogy with a theory in motion perception. The motion perception theory begins from the fact that humans see a group of isolated points as moving as a single object whenever some translation in three dimensions will account for their collective behaviour. This occurs despite the differences of the motions of the points in the two dimensional perceptual plane. It is a form of object shape constancy. The motion perception theory then notes that the two dimensional behaviour of the points may be divided up into the maximum common component and the residual differences. Following a few examples, this is not developed further; there is little indication of how these two parts can be used to determine the characteristics of either the object or its motion, but that is not the central topic of the paper.

The translation from motion perception into colour perception is the idea that the common colour component of the scene can be removed to eliminate the colour of the illumination, leaving the residuals

- 49 -

attributable to the object colours. Unfortunately, this notion neither corresponds to observations of human vision, nor is it a particularly effective procedure. McCann, Hall and Land [1977] have shown that the average colour of a scene can be unrelated to the colour of the illumination, and that, in ordinary scenes, factoring out the average flux colour can yield entirely wrong results. It is not entirely clear that it is the pixel-wise average used by McCann et. al. that Bergstrom intends to use, since no actual computational maximum is specified. As will be seen later, a region-wise average is usually slightly better. In essence, the issue is one of scaling the colour cube. Without stating it, the paper is proposing that the colour cube can be scaled by some form of average colour across the scene. This scaling is inappropriate for the many scenes which have an average of reflectances that is not grey. For example, a scene of a garden will have a lot of green objects, and often these will be a variety of shades. When the average is taken, the green will be attributed to the illumination, which will give a systematically wrong value for each reflectance.

The mechanism will also be fooled by a scene with different illumination areas, so that one area is mainly under one light, while another is mainly under another. The average will be controlled by the relative sizes of the two regions. It might be said that the two areas could be separated and two different compensations considered, but it is not clear how to do the separation.

The difficulties can be seen to result from the assumptions in the paper. It makes no attempt to go beyond the Mondrian world yet does not explicitly restrict the domain, and it uses Land's step edge detector. However, the paper does go beyond Land by not dismissing

the brightness changes as mere noise. Bergstrom points out that a scene-wide brightness gradient is due to the distance from the source, and this information can be extracted explicitly rather than discarded during edge detection. This observation is only strictly true for Mondrian worlds, but we will use a similar principle later.

The paper has some other contributions. It contains a concise summary of those previous colour theories coming from perceptual psychology. It is also an example of the application of the Intrinsic Images paradigm to colour perception, even before that framework was announced [Barrow & Tenenbaum, 1978]. In colour research, that paradigm can be seen back at least as far as Katz [1930].

Additional Retinex research was reported by Walker [1979]. He and Robert Szabo replicated much of Land's two-colour projection phenomena using ordinary non-specialist equipment. Full colour appeared even with extreme input filters, neither of which passed red or green. The exposure of the negatives while photographing the original scene was shown to be not critical. However, it is surprising that this early work was still of interest at such a late date, when so much remains to be explored in the new Retinex results.

In addition to Retinex researchers, there have been several new theorists on constancy, as there have been in previous ages. Flock [1984] has announced that surface reflectance can be determined for every region by comparing the matt reflectance of the surface, which is influenced by the reflectance, with the specular reflectance, which is not. He holds that this is always possible, because every surface is both partially matt and partially specular and because surfaces tend to curve and so pass the surface normal to present a

highlight to the eye. While we agree that highlights are usable for this purpose when they can be found, identified and measured, it is difficult to take seriously the claim that they are strong enough on every surface to be accurately measurable. He also does not notice that the specular flux includes the matt flux at that pixel, which should be subtracted off.

Buchsbaum and Goldstein [1980] have re-asserted the old notion that the illumination is blindly estimated from the retinal surround, as if in ignorance of Gilchrist's work and Land and McCann's many demonstrations that average flux and similar metrics are inadequate and inappropriate for constancy.

Richards [1978] has presented an interesting note on lightness constancy. In addition to the fact that the rods are more sensitive to direct flux, they are much more responsive to light hitting the retina from oblique angles, including intraocular scattering. He offers a mechanism whereby the difference between these is used to determine a "unique grey" reflectance of twenty percent. Although this theory too falls down because of its average flux postulate, it is certainly a novel mechanism to estimate average flux.

The approach to colour processing that has been discussed thus far concentrates on studying the characteristics of the biological vision system. This involves attempting to determine the internal mechanism of a very complex system, and also analysing what information in the image can be extracted to determine the reflectance of the objects that produced the scene. Actually building a colour imaging system is an entirely different process. No longer is it possible to work top down, studying the general behaviour first and then extracting layers of detail. Instead, the most detailed, and for the present

purposes uninteresting, parts must be built first. The development typically mimics phylogeny. Physical colour receptors, signal processors, and storage to hold the image during processing are initial requirements. When these are built, the next step is basic image processing, finding primitive characteristics such as gradients, edges and regions. It is only when a large collection of tools are available that work on high level processes such as constancy can be attempted. With this in mind we will now look at the range of colour image processing projects to date. Many of these projects ran concurrently with the present research. It is also to be noted that most of them are intended to be practical solutions to imaging problems such as sorting or object detection. In this case, the aim is to avoid constancy problems rather than to undertake additional research towards solving them.

A new research group beginning the study of colour often does not realise that constancy will be a problem. Ito [1973] was only interested in specifying the hardware tools that seemed necessary for colour vision. The particular issues to be studied were not considered, but there is the implicit assumption that extracting object characteristics directly from the flux is no great problem. Later [Ito, 1975] their research facility was in use, but with a substantially altered design. There was now an interactive facility, enabling single command transformations between different colour representations. Given controlled illumination, this allows distinguishing features of objects to be found visually, and manually programmed into a robot vision system. They explicitly limited themselves to easy industrial problems.

For industrial applications, the simplest solution is chosen not

just for economy and reliability, but because hardware is often outdated very quickly. Ueda, Matsuda and Sako [1980] needed colour detection in assembling and painting. Any equipment used in a painting environment must be robust and cheap enough to replace frequently. Their solution was to mount three photodiodes, each with a different colour filter, onto a robot arm. Although it is effectively a one pixel camera, it is well suited to this environment. Colour identification is done by table lookup in a taught-by-example database. Unlike most other examples of teaching by example, they use a single large database for all applications. However, their reported methods do not differ substantially from Kato [1975].

Among the first to research the general application of colour to machine perception were Yachida and Tsuji [1971]. Their scenes were limited to coloured blocks, viewed in an image sixty-four pixels square at fifty intensity levels per colour. Like many after them, they "normalise" the colours by dividing each by the sum of the three at the pixel. They considered three metrics for distance in colour space, choosing to use the component with the largest difference solely because it was faster, such was the limitation of the hardware. An important problem for them was control of attention, because the "data was overwhelming". Their methods required that the background reflectance must be known and uniform.

Jarvis [1982] has essentially repeated this work a decade later, while looking for a method quick enough for industrial viability. The essential difference is that the new work has use of a laser range finder, which instantly gives almost all the object boundaries because of the discontinuities in depth that occur between most

objects.

Even today, image processing groups are still rediscovering the usefulness of colour. Nguyen, Poulsen and Louis [1983] halved their five percent error rate for cervical cell classification by adding colour to their list of cell attributes. They were not the first to perform useful colour processing using only two image colours, as opposed to the traditional three.

Some colour methods define a function for combining the colour values into a single value per pixel. This is equivalent to changing the spectral response of the camera. This is not colour processing in the present sense, but essentially a monochrome camera with a response curve selected by software. However, it can be very useful in specific applications. Notice that among monochrome films, the panchromatic film, which has equal response across the spectrum, is usually considered less informative by humans than orthochromatic, which makes one end of the spectrum dark and the other light.

Most of the practical image processing systems in use today avoid the problems of colour and lightness constancy by controlling the illumination, either with the lighting as part of the imaging system, or by specifying that it must remain constant after the teaching phase. The system designed by Fiorini and Montevecchi [1982] is a box which fits on a conveyor belt, shielding out room light, and controlling the relative positioning of camera and fluorescent lighting. Discrimination is eight bits in each of three channels, and colour identification speeds in excess of ten objects per second have been obtained.

Inspectrum [Loughlin, 1982] is a commercial turnkey visual inspection system designed to synchronise with manufacturing

hardware. It is taught a set of inspection points, which are single pixels, and learns from examples which RGB values to expect. Thus it is dependent upon the lighting remaining constant, and also a fixed presentation of each object. When used with a CCD camera, which typically has malfunctioning pixels, it is necessary to carefully select the inspection pixels used in a system of this design.

Chen and Milgram [1982] discuss the Retinex theory, but still choose to control the illumination. A remarkable characteristic of their system is that it is binary in each of three colours, whereas most of their contemporaries use 256 levels per channel. This is possible through careful scene-specific choice of threshold. Colour separation is achieved with a monochrome camera by rotating a wheel of three colours if front of it.

Yoshimoto and Torige [1983] also use binary colour. Notice that binary colour provides exactly eight different flux values per pixel. A subtle but important distinction is that these can not be compared in the same way as eight-level grey-scale, because similar colours are not necessarily numerically adjacent values, as they are with grey levels.

A domain with fairly constant illumination, which has received more attention than any other is the processing of Landsat photographs, which have four or more colour separations, including infrared. Tou and Gonzalez [1974] were able to locate rivers, forests, green fields and crops, by training their program on samples from the image. They could not reliably distinguish between crops.

Rebollo and Escudero [1977] also learn from examples, but are better able to handle the ambiguity where cases overlap. They statistically minimise the intersection.

However, learning from examples proved inadequate, because variations in latitude, altitude and planting time defeat classic pattern recognition techniques. Wheeler and Misra [1980] found that the rate of colour change over a few days can identify the crop and also the point in its growth cycle. They also noticed that the information in their four-dimensional data of crop colours was essentially confined to a two-dimensional subspace of approximately brightness and greenness, and thus greatly reduced the computational complexity by transforming the co-ordinate system.

Addressing the same problem, Davis, Wang and Xie [1983] use probabilistic relaxation to achieve improvements in classification of agricultural crops in aerial photographs. A novel feature was that they compare photographs from different seasons. The colour change between the images specifies the crop and estimates its yield.

Schaerf and Mauer [1982] improved their Landsat classification by combining colour and texture, both of which had previously been used separately.

Whereas Landsat processing generally is done by simply classifying each pixel, other colour domains require that larger objects be found, which then requires either region or edge finding. Occular fundus photographs, which are colour pictures of the retina, are a very useful diagnostic for early detection of human circulatory disease. Falconer, Barrett and Kottler [1979] processed the red image independently for the gross features, and then found that the features could be distinguished by colour - vascular is red, retinal cells are green, and the optic nerve is blue. Like many other teams, they were surprised at the weakness of their blue image. This can usually be traced to a deliberate bias in general purpose monochrome

cameras. This bias that minimises blue makes monochrome pictures more pleasing to human observers. This group also failed to match the pigments of their input filters to the pigments of the colour photograph, thus increasing distortion while showing an ignorance of the dimensionality of their medium.

At the same time Akita and Kuga [1979] regarded their own attempts as unsuccessful due to the variety of colours and shapes. They were using a dynamic threshold set by trial and error for small blocks. They were the first to admit that the fundus images have shading and uneven lighting. Later [Akita and Kuga, 1982], they had changed from deterministic methods to relaxation and verification. The latter stage classifies unknown pixels after the relaxation has converged. They found that intensity distinguishes arteries and veins.

Relaxation is an image processing technique where each pixel is modified according to particular characteristics of its neighbours. The new image is then processed again in the same manner until the image comes to a rest, with no further changes possible. An example is when each pixel changes half way towards the average of its neighbours that are not across an edge. This program would eventually converge to a constant value per region.

Eklundh, Yamamoto and Rosenfeld [1980] experimentally found that relaxation gave better error reduction in multispectral pictures than did the pre- or post-processing methods that they tried.

Davis and Rosenfeld [1981] observed that most pixels are internal to their region, and so all the neighbours are in the same colour reflectance class. Edge pixels, which are less frequent, have neighbours of two classes. From this they developed cooperative classification. Their method assumes that the flux is flat or

locally planar.

Kittler and Foglein [1984] used a recursive contextual statistical decision process very similar to relaxation for colour classification. They present statistical proofs backed up with experiments on artificial and some real images. However, even with an understanding of statistical methods it is difficult to determine the model of the scene domain used in this branch of image processing. In this and many similar studies, the objects of interest are the algorithms, rather than the scenes. The actual topics addressed may be convergence of the relaxation or choice of stochastic model. This is complementary to the body of work which looks for constraints within scene images and methods of utilising them in scene analysis.

One alternative to relaxation is cluster finding. This approach observes that flux values for the same region tend to group together within colour space. Ohlander [1975] presented a new way to separate these clusters, which usually overlap. It begins with the whole image and recursively descends, at each step looking for the most appropriate feature to divide regions. In practice he used hue and intensity as the main features.

Price [1977] refined Ohlander's method by adding a planning stage, which improved the speed by a factor of ten. He also added some multi-spectral texture measures. The project was to detect changes between two images of the same scene, and eventually, object colour became just one feature, together with size, shape, position, and so on. An enhanced form of their combined colour segmentation method has been reported [Ohlander, Price and Reddy, 1978].

Connah and Fishbourne [1981] discovered the same problem of

overlapping clusters in their industrial implementation. Without the benefit of the Ohlander-Price algorithm, they created comparable ad hoc methods. Clusters are detected by finding local peaks. Then clusters which have too great colour variance are broken up, causing a secondary region finding process. Initial colour edge finding is by a modified Sobel operator. The Sobel was an early edge finder which finds only horizontal and vertical edges. They modified it to also include diagonal edges, but this did not improve the edge catchment. They also observed mixed pixels, which were unexpected and ignored. A mixed pixel occurs whenever the small area viewed by a pixel of the camera includes an edge. It thus takes on a value between that of the two regions, roughly in proportion to the visible portions of each.

Ohta, Kanade and Sakai [1980] made a systematic study of a range of colour functions for region segmentation. Using their hardware and choice of segmentation algorithm, their best three, in order were:

$$(R+G+B)/3, \qquad R-B \qquad \text{and} \qquad (2G-R-B)/2$$

Kender [1976, 1977] noticed that for most colour coordinate systems, non-linear transformations do not preserve uniform distributions of digitised images. This implies errors in almost every normalisation and distance metric, and he provides colour spaces where these problems do not occur.

A region segmentation method not unrelated to Ohlander was found by Holla [1982] while modelling a particular theory of early visual processing. The model makes no attempt at colour constancy, but computes red-green and yellow-blue channels, then looks for features in the two dimensional histogram.

- 60 -

The use of colour as an object feature is a frequent theme. Tenenbaum [1973] made a landmark suggestion when he proposed moving beyond the sterile and monochrome "blocks world" which had previously been used in scene understanding research. He made plain the assumption that the scene was rich with redundant information. Some previous researchers had treated the problem as if there was very little usable data. This eventually led to the Intrinsic Images philosophy [Barrow and Tenenbaum, 1978]. However, this early paper was interested in pragmatic identification of specific objects in an office environment, rather than general scene understanding. It also assumes that it is trivial to find object colour.

At the next stage, Tenenbaum, Garvey, Weyl and Wolf [1974] were defining complicated transformations between the red, green and blue space, RGB, and the intensity, hue and saturation space, IHS. Then Kunii, Weyl and Tenenbaum [1974] defined a region as an area having constant flux colour, and proposed finding the colour by a gross colour match followed by iterative comparison.

Later [Tenenbaum and Weyl, 1975], choice of colour function for region merging had become a primary area of research. They found that equal-brightness colour was less useful than non-colour brightness. This is not surprising to perceptual psychologists, who have long noted the perceptual difficulties of isoluminance borders. After comparing five measures of boundary and region contrast, they concluded that the maximum of the two was the best measure tried. None the less, it was always the case in their experiments that erroneous merges appeared long before completion of correct merging, so their system was made semi-manual.

A number of results in colour processing have implications for

colour constancy. Hall and Andrews [1978] have shown how a three
colour picture can be compressed into one bit per pixel with no
visible degradations on reconstruction. A perceptual power spectrum
puts the noise where one does not see it. This either shows the
extreme redundancy in colour images, or demonstrates the power of the
eye in processing limited information. In a related field, work has
just begun on the first use of colour in multispectral image
restoration [Hunt and Kubler, 1984].

Separating out Intensity is in fact a type of colour constancy,
because this removes the effects of the amount of light available at
a scene point. However, it can only work if there is a single
constant colour of illumination and negligible secondary reflectance.
Further, it loses a dimension of information, so that objects of
similar colour but different lightness are confused. Ali, Martin and
Aggraval [1979] approximately factor out the intensity and then
perform blob detection on the remaining two dimensions, which happen
to approximate red and green.

Although we accept the value of "normalising" the colour at a
point, by factoring out the intensity, we will now show that the
universally accepted method of doing this is incorrect, and a correct
method will be presented. A standard text on Image Processing, Duda
and Hart [1973] notes that normalising the red, green and blue inputs
removes the effects of local changes of available light, and leaves
only two dimensions of colour. However, neither they nor their
successors have noticed that the normalisation must be the colour of
the illuminant. The universally used formula is the sum of the three
colour values, red, green and blue. It is only by chance that this
sum approximates white. In most cases the camera response through

the coloured filters is far from equal in these three components. Compound this with the fact that artificial light does not have a flat spectrum, and that normal television cameras have a deliberately skewed spectral response. This single, seductive mistake can be seen to cause a substantial amount of the scattering of pixels in colour space in much of the reported work. In the typical controlled environment it is not difficult to overcome this problem. Test samples may be used to determine the response of the illumination-filter-camera combination, and scalars chosen so that the Intensity function matches the change in brightness.

It is easier to see the problem by an example, but first it should be noted that the problem with this intensity metric is when comparing pixels from different reflectances. If the two pixels are from the same surface region then this metric, or any other, will work, given the ideal conditions of no noise and a single colour of illumination, and no secondary reflection or background illumination. Say two pixels are from the same region but the second gets half as much illumination. Then each channel will receive half as much illumination, so any linear combination of the channels will give half as much illumination in the second as the first pixel. Therefore, when it is normalised by half the amount, its values are doubled to exactly equal those of the full illuminance pixel. This may be part of the reason that this metric is so easily accepted. But notice that if the normalising metric were simply the first channel, say red, then the same property would still hold.

Whenever pixels from different regions are compared the metric fails. Consider an example. Say that the receptivity of the camera-filter-illuminant combination is in a ratio of 4:5:2 for the

three channels viewing a white card. Two objects have channel-wise reflectance triples of $(.5 \ .8 \ .6)$ and $(.2 \ .4 \ .4)$, respectively. To reduce the variables of the example, they may have the same amount of illumination. As justified in the next chapter, the flux is the component-wise product of the illumination and reflectance; if there are 4 units of light and a reflectance of .5, then 2.0 is resulting flux.

$$\begin{array}{ccc} \text{Illum} & \text{Reflect} & \text{Flux} \\ (4 \ 5 \ 2) \times (.5 \ .8 \ .6) & = & (2.0 \ 4.0 \ 1.2) \\ (4 \ 5 \ 2) \times (.2 \ .4 \ .4) & = & (0.8 \ 2.0 \ 0.8) \end{array}$$

Using the metric R+G+B, the "intensities" are

$$2.0 + 4.0 + 1.2 = 7.2$$
$$0.8 + 2.0 + 0.8 = 3.6$$

This is immediately seen to be wrong, since it says that the second pixel received half as much light, when in fact they received exactly the same amount. The normalisation brings the two intensities to the same level, which can be done by doubling each channel of the second to finish with

$$(2.0 \ 4.0 \ 1.2) \quad \text{first pixel}$$
$$(1.6 \ 4.0 \ 1.6) \quad \text{second pixel}$$

This says that in the first channel the second pixel has eighty percent of the reflectance of the $first$, when it is actually forty percent. In the second channel the two seem to have identical reflectance, whereas the first is double. In the third channel the second pixel appears to reflect more, whereas it actually reflects less. Had the brightness of the illumination also varied in this example, as it does in real scenes, the numbers would have been some other set of arbitrary incorrect values.

A decade after Duda and Hart, Ballard and Brown's textbook [1982] claims to appreciate the problems of understanding colour vision in man and machine. However, it is only interested in transformations between RGB and IHS systems. It still states the "fact" that intensity is R+G+B. Finally, it cites Land [1977] only for the "startling" phenomenon of two-dimensional colour.

Cohen and Feigenbaum's Handbook of Artificial Intelligence [1982] is similar, giving good treatment to classic colour systems, as well as two historic methods of colour edge detection that come from the techniques of Pattern Recognition rather than modern Image Understanding. They do not cite any reference which deals with

compensating for the effects of the illumination. Although they give the standard incorrect formula for normalising colour vectors, no reason is given for its use.

Prior to the present study, there have been several computer implementations of Retinex theory. Mayhew and Frisby [Frisby, 1979, pages 135-138] have implemented Marr's [1974] retina. By modelling the neurology, they have replicated simultaneous brightness contrast, the Craik-Cornsweet-O'Brien illusion and Mach bands, and were going to replicate Kanizsa's Triangle, where illusory contours produce illusory brightness. It is significant that the computer model has the same perception as humans, even under illusory stimulus.

The paradigm for the present work has evolved in subtle but important ways. During the initial stages the intention was to utilise colour cues in a manner analogous to Waltz [1972] processing, in order to achieve constancy. "Waltz filtering", as it is called, processed line drawings of scenes in a way that found the correct three-dimensional meaning of those scenes. It did this by classifying each local feature - corners, edges and shadows, and then gradually eliminating all inconsistent possibilities. Such inconsistencies were generally two junctions giving different meaning to the same line which connects them, such as one junction labeling the line as a shadow while the other labels it as a region boundary. The remarkable thing was that when all the inconsistencies were removed, there was almost always a unique interpretation remaining. It remains possible that such processing of colour edges, attributing to them a certain amount of reflectance change and a certain amount of illumination change, will have a similar result. However, that

hypothesis has largely been overtaken by the present study.

The paradigm used at the second stage of the present work was that of the Intrinsic Images, by Barrow and Tenenbaum [1978]. This paper is in two parts. The first sets out the general philosophy and the second is a specific contribution to achieving the processing goals of the first part. The philosophy is that there are a collection of attributes at each image point, including surface normal, distance from viewer, incident illumination, surface lightness, surface colour and so on. They represented these as different parallel planes, and showed that the planes were interrelated in specific ways both between planes and between cells in a single plane. They pointed out that within most feature planes, values usually change smoothly, except at lines of discontinuity, and that these lines usually have special perceptual meaning and also effects in other planes. For example, a discontinuity in the incident illumination is a shadow. A discontinuity in distance is an occlusion, and this is usually associated with a change in the reflectance plane. Their paper suggests that perhaps there is considerable redundancy in visual images, whereas previous workers in computational vision had generally implied that information is scarce.

Despite the usefulness of their philosophy, the second part of the paper contains an important mistake which has not been reported elsewhere. One significant contribution which they make is to provide a way of distinguishing which side of an occluding edge represents the object which is closer to the viewer. They then use this edge test in a world of even background light and one point source to classify all the occlusions in the scene and complete their processing. The edge test requires that the near object be

illuminated by the point source. Now, all of the possible surface normals are representable by a sphere. And so the equator as we look down on a pole is the set of all possible occlusions. If the sphere is illuminated by a point source, only half of the sphere will be in light. Thus no more than half of the occlusion edges in their scene could be measured by this edge test, and their algorithm can not proceed due to lack of information. It is surprising that this fault did not become apparent in their implementation. The problem can be rectified by having multiple point sources, each of a different colour, and using the techniques of the following chapters to separate the lights, which collectively illuminate most of the occlusions.

From Intrinsic Images, we have gradually evolved into a third paradigm, the Computational Vision philosophy expressed so eloquently by the late David Marr [1982]. This transfers emphasis from the act of processing an image to the study of the information within the image which can be extracted, and the meaning of the information. It insists that any vision theory be supportable by a computable method for implementing the theory, and in this it differs from some other expressions of "ecological" perception. However, it abhors ad hoc processing which looks to find an arbitrary method of processing the present image, without identifying the image attribute which relates to the feature being sought.

In keeping with this philosophy, the subsequent chapters distinguish particular attributes in colour images which are a measure of the colours of the illuminations and surfaces. The important parts of the theory have been supported by being implemented in working computer programs. Each program is described

at the end of the relevant chapter, but as much irrelevant detail is omitted as possible. The building and testing of the programs invariably lead to new insights into the issues of extracting colour attributes, and each of these insights is explained, although some of them are unanswered questions.

We now introduce an important principle in colour and lightness constancy which may be called the minimum number of sources principle. As was shown by the example in the preceding chapter, where both photographs and transparencies give perceptions of reflectance and lighting changes while the former has only reflectance changes and the latter has only illumination changes, changes in flux can in general be attributable to either changes in illumination or changes in reflectance. This is the essential problem of colour and lightness constancy. A given computational mechanism will normally disambiguate much of the illumination and reflectance within the scene, but typically there will remain some fundamental ambiguities, where an edge or gradient can be reasonably explained as either a surface or a lighting phenomenon. There are two parts to the principle of minimum number of sources. The first says that if the changes are in the colour of the flux, then assume they are caused by changes of reflectance in the absence of contradictory evidence. This is because there are less sources than objects. In the sense of the present discussion, which will be clarified below, a colour change must be due to either an additional object or an additional illumination source, and the latter is less likely. The opposite is true for brightness changes, since regions often have nearly constant reflectivity, while source brightness constantly varies, due to

distance from sources, shadows and angle to sources. This is the second part of the principle, stating that brightness changes are due to changes in illumination strength unless shown otherwise.

The difference between brightness and colour changes is generally apparent in scenes illuminated by a single source colour. Brightness changes are those where each receptor channel changes by the same ratio. However, with two or more colours of sources, the difference is not so apparent. Even in a Mondrian world, a red light from one side and a blue light from the other will produce a flux gradient that is seems to be a colour change at every pixel. If it is possible to detect this state, or to separate the effects of the different sources, then it too can be correctly identified as a change of brightness. In fact, chapter three gives a method of determining the number of sources in a scene region, and then a method of separating their effects.

Gilchrist, Land and the other Retinex researchers use rich scenes for perceptual experiments, unlike many others. These are more likely to give information about the system's abilities, rather than about how it breaks down under degenerate stimuli. Considerably more research is possible in this area, and in chapter seven several classes of experiments using less degenerate scenes are suggested.

Chapter Three

Mathematics of Colour Images
Region Finding in Flux Space
Determining Relative Surface Colour

The following discussion assumes an idealised world in which to perform colour constancy computations. We begin by examining the characteristics of this world and the different types of colour information which may be extracted from it.

Each point in the image, called a pixel, is assigned a scalar value by a monochrome imaging mechanism. A polychrome imager, whether it is camera or eye, will distinguish different colours at each pixel, and so produces a vector of scalars. It does this by producing a scalar value for each of a small number of receptors, with each receptor being sensitive to a different band of wavelengths. In the eye there are believed to be three different types of cone receptors [Marks, et. al., 1964], each with a different receptivity curve, plus the rod receptors, which have a further different receptivity curve. In a typical colour electronic camera, such as those used for television, there will also be three colours of receptor, roughly associated with red, green and blue. A monochrome camera can be made to receive colour information by taking a sequence of photographs of a single scene, each through a different colour filter.

The information received from each of these sources is equivalent for our purposes. For a given point in the image there may be for example 12 units of "red", 10 units of "green", 7 units of "blue", and 11 units of "grey". But notice that the "red" receptor does not

receive only a single wavelength corresponding to a particular red. Instead, it is a broad band receptor which is most sensitive to the wavelengths that happen to bear the same name. For the present purposes it is more useful to think of colour in terms of the ordered list of values 12, 10, 7, 11 than in terms of any particular colour names.

We will refer to this as the colour vector. As will be seen throughout the following chapters, the analogy with spatial vectors is not perfect. In particular, the most common arithmetic operations between vectors are component-wise. The arithmetic implicitly assumes that receptor response is linear. That is, if the illumination which produced the above vector was doubled, then each component of the vector would also double, to become 24, 20, 14, 22. Many actual cameras have an approximately exponential response, but for our purposes it is assumed that the response curve can be measured and compensated for in pre-processing.

The vector model is different from, but not incompatible with, a Newtonian view which can be described from two viewpoints. First, there is the spectrum, and the observations from it such as that a wavelength of 450 microns is seen to be red. However, noting that each visible wavelength seems a different colour is not the same as showing that every colour can be characterised by a wavelength. The most obvious problems are the purples and browns which are not represented in the spectrum. In addition almost every naturally occurring colour is richer than the spectral colour which approximates it. If a prism is used to determine the spectrum of a natural pigment, most or all of the spectral wavelengths will be present, but some will be brighter than others. Essentially, a

natural colour is not a single number representing wavelength, but an intensity function over the entire domain of visible wavelengths.

As an aside, note that artists long ago observed this problem, and attempted to remedy it by representing colour as having the three components of hue, brightness and saturation. Hue was essentially wavelength, which we can take to be the highest intensity of the spectral function. Brightness was overall flux, which is equivalent to the integral over the spectral function. Saturation was the "purity" of the colour, largely an expression of how intense were the wavelengths which were distant from the hue value. It would be hard to make a computational equivalent to saturation, because it was defined only in terms of human perception, which in turn is defined by the cone and rod receptors of the eye.

The second and more complete Newtonian view is that in order to fully record a colour it is necessary to record its complete spectral function. For an arbitrary colour this would require measuring and retaining the intensity of each discernible wavelength, which would require a large or infinite number of receptors of each pixel. Fortunately this problem is reduced by two typical properties of natural scenes. Firstly, natural spectrum intensity functions tend to be relatively smooth, so that if the strengths of two close wavelengths are known, then the values of the wavelengths between them can be reasonably approximated. Secondly, a natural pigment is often bright in more than one region of the spectrum. Thus, if pigment A is being added to a mixture of pigments, and it is bright in both yellow and blue, then the intensity spectrum of the mixture will be increasing in both these colours. This is more likely to be noticed by a small set of receptors than a change of a single

wavelength.

Although a large or infinite number of receptors is needed to record the full spectrum of a colour, Land has shown that it seems possible to record the human perception of the colour using only four receptors. He does this with his "Retinex" filters which are closely matched to the receptivity of the four human channels. At this level of approximation, he can ignore the small variations between individuals. For a given colour vector generated by Retinex filters there is one colour perception. However, that vector may have been generated by any of a large range of spectra, but each of those spectra would be seen the same by a human observer.

Returning to practical imaging devices, it can be seen that using broad band receptors is not just a convenience for manufacturing and a means of allowing through a more measurable amount of light, but also a benefit for measuring diverse natural colours. In general it is best if the entire useful spectrum is covered by the receptors, which of course will overlap. It is also better to have as many receptor types as possible, given that each is independent. In a restricted environment containing only a few pigments, such as an assembly line, a more specific choice of receptor spectral curves may be optimal.

In any application it is essential that the receptor curves be mutually independent. If one curve was the sum of two others then its value would always be calculable from the other two, and so it would never provide any additional information. However, a brightness or "grey" receptor, such as in the first example, would not necessarily be dependent on the colour receptors, as might naively be assumed. The fallacious logic is that each colour

receptor responds to brightness changes as well as colour shifts, and if all the colour receptors increase by the same amount, it is probably due to an overall brightness change. While that deduction may frequently be true, it is unlikely that the hidden assumption, that a "grey" receptor has the same sensitivity as the direct sum of the colour receptors, is true. A grey receptor would be a more or less flat curve across the spectrum, and would pick up those points "between" other receptors, which is where the sensitivity of other receptors was low. A brightness receptor is thus not necessarily the same as the sum of the other receptor values. For the purpose of measuring and distinguishing spectrum intensity functions it can be treated as just another receptor.

There is some issue in the literature about the role of the cones in colour perception. Many researchers, including Land [1977], believe that the space of all perceivable colours is three dimensional, and one would assume that the three types of cones span this space. Yet on the other hand Land's laboratory at Polaroid [McCann and Benton, 1969] has shown that nearly the whole range of colour perceptions can be produced from an image which is only perceived by the rods and the long-wave cones. Finally, there is general agreement that at normal daylight levels the rods are saturated and therefore useless.

We now consider a way of resolving this apparently contradictory data, which also gives some insight into colour mathematics. Assume that colour space is three dimensional and behaves according to Land's observations. In particular his illustrations [Land, 1977, page 120] show that the naturally occurring colours fall near the main diagonal. This diagonal represents the grey range from black in

one corner to white in the other. The values tend to fall within a narrow double-cone shape which has its points at white and black and its widest cross section near the middle of the cube, as shown in diagram 3.1. Now consider the hypothesis that the rods also index the cube, on an axis which is roughly along this main diagonal. The
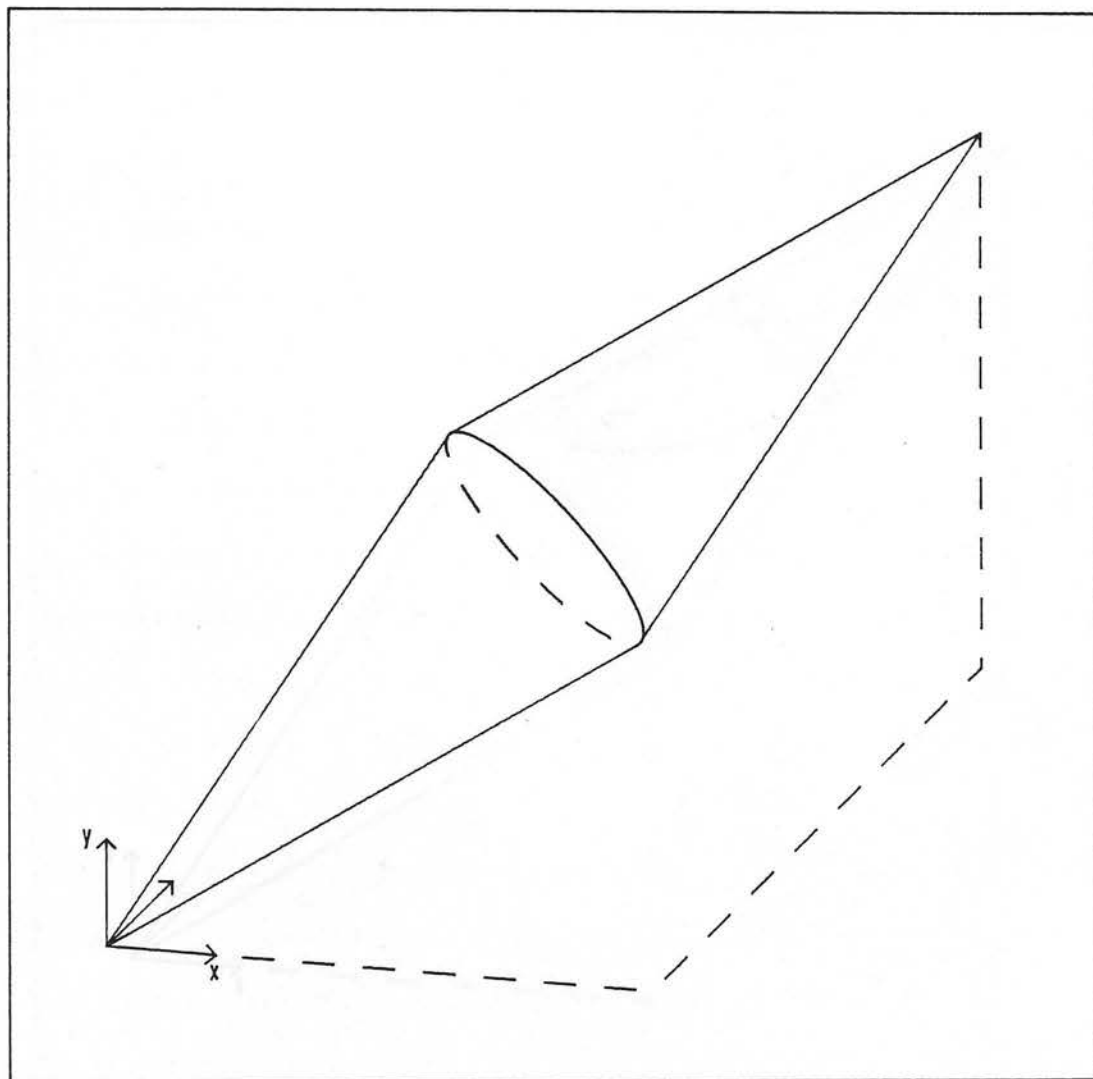
3.1 Normal Reflectances lie within a Double Cone

dimension scaled by the rods need not be independent; it can be redundant information that refines the cone measurements, or supplements them when the light is dim. Now in Land and Benton's experiment, only the rods and long wave cones are stimulated, but

these are two independent dimensions. Specifying a value for each limits the perceived colour to the intersection of two planes. This line is further limited to the small range of colours in the cross section of the active diagonal of the colour cube. Therefore the colour is already very nearly specified, as can be seen by the
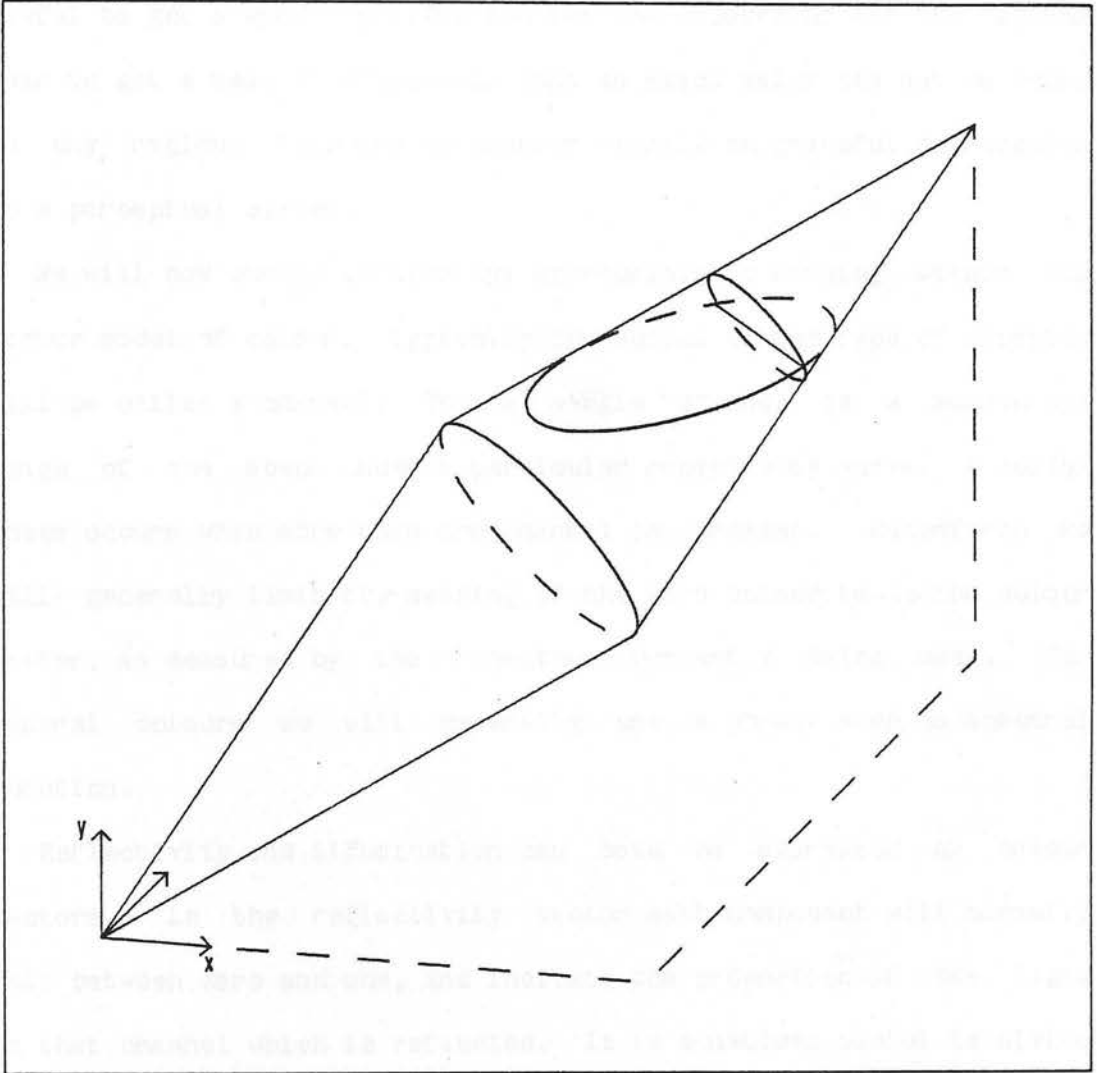
3.2 Intersection of Rods and Long-Wave Cones

intersection in diagram 3.2. Finally, the vision mechanism is set up to see only one colour per object; there are no common examples of something appearing to be both red and green at the same pixel at the same time. So the colour constancy mechanism must specify exactly

- 77 -

one colour to the brain as the colour of the cell. The only surprise is that no "error" signal has been returned as well, as when the region may appear shimmering or constantly changing its appearance between the possible colours. On the other hand, the entire scene is being perceived under the same reduced-light conditions. It is more useful to get a good approximation for the colours of all the regions than to get a mass of complaints that an exact value can not be found in any region. This may be another example of graceful degradation in a perceptual system.

We will now choose terminology appropriate to working within the vector model of colour. Typically the output of one type of receptor will be called a channel. Thus a single channel is a monochrome image of the scene under a particular receptivity curve. A colour image occurs when more than one channel is present. Henceforth we will generally limit the meaning of the word colour to be the colour vector, as measured by the receptors currently being used. For natural colours we will generally use a phrase such as spectral function.

Reflectivity and illumination can both be expressed as colour vectors. In the reflectivity vector each component will normally fall between zero and one, and indicate the proportion of the light in that channel which is reflected. It is sometimes useful to divide up the illumination or flux vector into a normalised colour and a brightness scalar. Various normalisations are appropriate to different computations. One is that the sum of the components in the normalised vector is unity. Another is that some particular component is always one. The word white will often be used for a surface of strong neutral reflectivity, such as $(.5, .5, .5)$. Grey

will mean a weaker neutral reflectivity, such as (.12, .12, .12).

We also assume that surfaces are Lambertian, that is that they accept and emit light equally in all directions. An important axiom is that   the equation

$$F = I * R$$

holds as a linear component-wise product. For example reflectance (.5, .3, .4) under illumination (8, 12, 13) gives a measurable flux vector of (4, 3.6, 5.2). Land implicitly relies on this property in his work, and so we use it here.[1] However, it is not clear that it is necessary; although it is useful in computations, each result is also seen in terms of an intuitively reasonable method of extracting commonly-occurring data from a scene.

In fact, a less restrictive axiom seems to be adequate. It states simply that brightness is a scalar; if twice as much illumination, reflectance or flux is present, then the measurable value in each component is doubled. This then means that brightness changes in both illumination and reflectance show up equally in the flux, or that illumination and reflectance brightnesses are interchangeable. This is equivalent to

$$f = i * r$$

for the brightness scalars of the previous vectors, after normalisation. This is equivalent to saying that, for any given surface and sources, there exists a constant vector C, such that

$$F = C * I * R$$

It is never necessary to compute this conversion vector, which only serves to re-scale the quantities. The essential step in colour constancy is to separate the effects of the illumination and reflectance. Yet R must always be measured with respect to some

1. It is mathematically true when the receptor functions are narrow in comparison with the wavelength-intensity variations in reflectances and illuminations. The required property is that the illumination and receptor (over)

- 79 -

illumination; if there is no illumination, then the flux is black, which gives no information. The actual question in colour constancy is not "What is the reflectance?", but "What is the flux under the standard illumination?", where standard in diurnal animals is probably direct sunlight. That is, it is not necessary or practical to determine either the absolute reflectance or illumination independent of the receptor system. In any event, the central concern of the present research is to specify the environmental clues which allow colour constancy, rather than to provide the theoretical foundations of colour perception.

In practice many implementations use the logarithm of the flux in their computations. This makes

$$f = i + r \qquad \text{(logarithmic values)}$$

which is quicker than multiplication on most machines, and is regarded by many to be closer to biological computational methods.

In many computations we will ignore the effects of secondary reflection and background light, and these will be treated explicitly in chapter six. The most common example of secondary reflectance is mutual reflectance, where two surfaces form a convex corner, and so some of the flux from each surface falls on the other and provides a bit of additional illumination. The problem is that this new illumination is of a different colour from the primary source, and so complicates the flux. For the moment we will take it that these effects are negligibly small and can be ignored. In chapter six they will be regarded as a useful source of information and handled by explicit processing. In the case of background, if it is of one consistent colour, such as skylight, then it usually may be treated as an ordinary light source. In this chapter there are no physical

(continued) functions are such that the reflectance term is effectively independent of the illumination.

requirements on the source, such as limiting it to be a distant point. We freely permit shadows, and they may be focussed or fuzzy. It will be found that the algorithm is comparatively robust with respect to accidental alignments in the scene, such as a shadow occurring along a reflectance edge.

This world initially has no noise in the imaging system, and no mixed pixels or sub-pixel-sized image regions. A mixed pixel is one where the pixel receptor receives flux from more than one region, and so usually gives a value between that produced by either region alone. Extensions to the algorithm to process mixed pixels and limited noise are not difficult, as will be seen in chapter seven.

We will assume, unless otherwise stated, that there are no sources visible in the scene, and so no light emitting objects. This includes fluorescent surfaces, which can emit more light at a given frequency than they receive. It will be seen that if one of the illumination sources is visible in the scene, it will be readily noticed. Since one of the critical steps in colour constancy is to determine the colour of the light sources, to have the source directly in view makes its colour directly known, and so reduces the work of the algorithm.

In notation, upper case letters will be used for colour vectors, typically F for flux, I for illumination and R for reflectance. Lower case letters, f, i and r, will represent scalars, either lightness or a component of the colour vector according to context. Thus in a four channel image, the components of the reflectance might be expressed as

$$R = (r1, r2, r3, r4).$$

The operation of component-wise multiplication between vectors will

be represented by the star symbol, *, or because of its importance in this theory, simply by juxtaposition of vectors. The symbols ^ and ** are used for scalar exponentiation. Very often the illumination vector I will be separated into its brightness scalar and its normalised colour vector, such as

$$I = i * I'$$
$$= i (i1, i2, i3, i4), \quad \text{where } i1 + i2 + i3 + i4 = 1.$$

When several illuminants reach the same point in the scene, the flux is the reflectance times the sum of the illuminations. For example, with three source colours the relation is

$$F = R * (i1I1 + i2I2 + i3I3)$$

However, if there are multiple sources of the same colour, then the colour vectors can be combined and only the brightness scalars are different. In the previous example, if sources 1 and 2 had the same colour, then the expression reduces to

$$F = R * ((i1 + i2)I1 + i3I3)$$

The brightness scalars incorporate all the effects which change the overall brightness of the illumination. These include distance from the source, angle between the source and the surface normal, and partial shading due to unfocused shadows, as well as the actual brightness of the source or sources.

We now begin the process of region detection, and start with an example. Region A has reflectance of $(.5, .2, .3)$, region B has reflectance of $(.4, .3, .5)$. Illumination is of a single colour, $(.3, .4, .3)$, so any illumination will be a scalar multiple of this vector. In a realistic scene, with non-planar objects, the intensity of illumination at points in a single region may vary considerably, for the reasons given above. The table below gives some sample

values for the fluxes Af and Bf, from regions A and B respectively, for different values of the scalar i, as well as the related values of incident illumination, I.

i	I			Af			Bf		
.5	.15	.2	.15	.075	.04	.45	.06	.06	.075
1	.3	.4	.3	.15	.08	.09	.12	.12	.15
2	.6	.8	.6	.3	.16	.18	.24	.24	.3
5	1.5	2	1.5	.75	.4	.45	.6	.6	.75

We now define the concept of flux space. Quite simply, it is the non-negative portion of a Cartesian space of the same number of dimensions as the number of receptor channels in use. In our present example this is familiar 3-space, with only the positive octant and its three faces in use. Each of the three channels is assigned to an axis, and each pixel in the scene is plotted as a single point in flux space. It turns out that relationships obscured in the picture of the scene become obvious in the flux space.

In the present example, all of the flux points in region A will fall on the same line through the origin. Similarly, all the flux points in region B will fall onto a different line through the origin. This immediately concludes region finding in this case. Every pixel which is not black will fall on one of the lines, and so is in that region.

When there is only one source colour, and at least two receptor channels, there are only two difficulties with flux-space based region detection. First, two regions which are physically separated may both have the same reflectivity, and so fall on the same line. Thus it is necessary to take into account the spatial arrangement in the scene. In this case this can be done after the fact, by checking for continuity in the resulting region map. We will find cases later

where it is better to take this information into account during the region finding. On the other hand, this "problem" may be regarded as a virtue, since all the disjoint regions of the scene which have the same reflectivity, and so probably the same material, are immediately found and grouped together, even though some may receive far less light than others.

The second difficulty is more problematic. Consider in the above example a third region, C, of reflectivity $(.25, .1, .15)$, which happens to be one half of vector A. All of its flux points will fall onto the same line as vector A. Using only the flux space technique, one might assume that this was the same material as A, except that it received only half as much light for some reason. There are three solutions to this. As we shall see throughout this thesis, in the general colour constancy algorithm, some examination of edges usually needs to be done to provide brightness constancy information. This topic is treated later. A second solution is to add more channels. These co-incidences should be less likely as more independent receptor curves are used, just as identical reflectances to humans often appear very different under infra-red imaging. A third solution is peculiar to controlled environments. If such co-incidental alignments are found in a sorting or manufacturing environment, then different filters could be fitted to the existing channels to remove the alignment. This would work unless the materials were precisely the same pigment, with one darker than the other. This is made less likely by the fact that the colours of white and grey pigments are rarely truly flat, and so their amount can theoretically be measured. Should there still be problems in this area, it may be worth altering the colour of one of the

difficult objects.

Next we consider a way to determine relative surface colours. What is the result of the component-wise ratio of an arbitrary cell in region A and an arbitrary cell of region B? Take three examples, Af when i is 1 divided by Bf when i is 5, Af when i is 2 divided by Bf when i is .5, and Af and Bf when i is 1 for both.

$$
\begin{array}{lll}
Af & Bf & \\
(.15\ .08\ .09) \ / \ (.6\ .6\ .75) & = & (.25\ .13\_\ .12) \\
(.3\ .16\ .18) \ / \ (.06\ .06\ .075) & = & (5\quad 2.6\_\ 2.4) \\
(.15\ .08\ .09) \ / \ (.12\ .12\ .15) & = & (1.25\ .6\_\ .6)
\end{array}
$$

The resulting vector is alway the same, up to a scalar multiple. In particular, the vector is a multiple of the ratio of the two surfaces.

$$A \ / \ B \ = \ (.5\ .2\ .3) \ / \ (.4\ .3\ .5) \ = \ (1.25\ .6\_\ .6)$$

It is quite easy to demonstrate that this will always be the case, by separating each flux term into its components of reflectance vector, brightness scalar, and illumination vector, giving

$$Af \ / \ Bf \ = \ A\ i1\ I \ / \ B\ i2\ I \ = \ i1/i2\ A/B$$

In a world of one illumination colour, the component-wise ratio thus gives the relative surface colours, provided that both regions are under that same source. This means that the ratio is independent of the colour of the source, and that independence is an important step in colour constancy. Note that in our third example of the ratio of Af to Bf, the flux ratio is the same as the surface reflectance ratio. This will hold exactly when the two pixels receive the same amount of light, and so $i1/i2 = 1$. That might happen for instance at the edge between two co-planar surfaces. On the other hand, the brightness ratio need not be unity for it to be measurable, but discussion of brightness ratios can be deferred.

When there are two illumination sources, the component-wise ratio between pixels does not have the desired effect. Consider an example. Take reflectances D = (.2 .5 .4) and E = (.4 .1 .3), and two illuminations, I = (10 12 14) and J = (15 9 6). Two such illuminations are shown in diagram 3.3. For a few combinations of
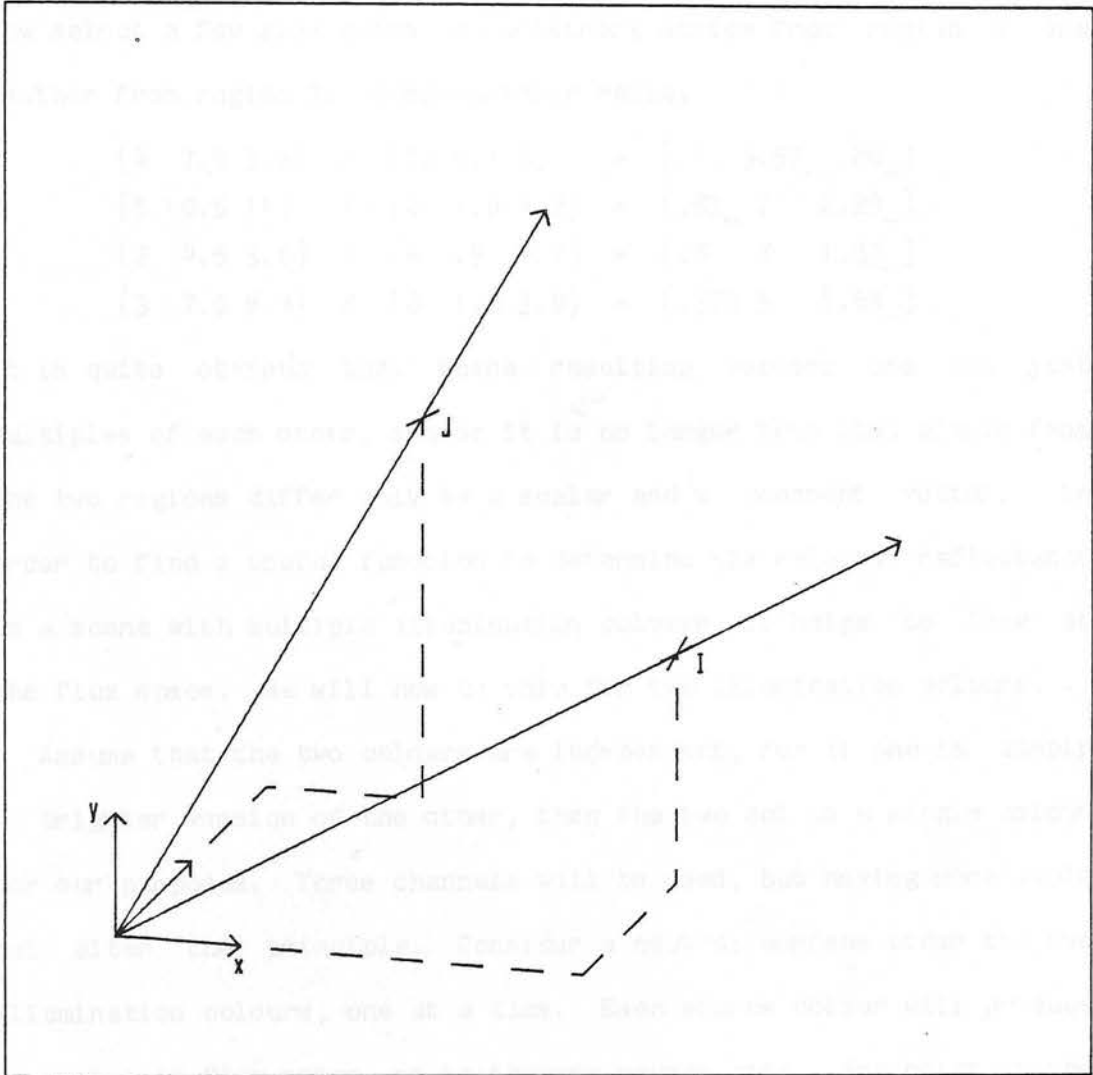
3.3 Two Illumination Vectors

the brightness scalars i and j, compute the flux vectors of those two

illuminations on the regions D and E.

| i	j	Df	Ef
1	1	$(5\ 10.5\ 8\ )$	$(10\ 2.1\ 6)$
1/2	1	$(4\ \ 7.5\ 5.2)$	$(8\ \ 1.5\ 3.9)$
1	1/3	$(3\ \ 7.5\ 6.4)$	$(6\ \ 1.5\ 4.8)$
1/2	1/3	$(2\ \ 4.5\ 3.6)$	$(4\ \ .9\ \ 2.7)$

Now select a few flux pairs, an arbitrary choice from region D and another from region E. Compute their ratio.

$$(4\ \ 7.5\ 5.2)\ /\ (10\ 2.1\ 6)\quad =\ (.4\ \ 3.57\_\ .86\_)$$
$$(5\ 10.5\ 11)\ /\ (6\ \ 1.5\ 4.8)\quad =\ (.83\_\ 7\quad 2.29\_)$$
$$(2\ \ 4.5\ 3.6)\ /\ (4\ \ .9\ 2.7)\quad =\ (.5\ \ 5\quad 1.33\_)$$
$$(3\ \ 7.5\ 6.4)\ /\ (8\ \ 1.5\ 3.9)\quad =\ (.375\ 5\quad 1.64\_)$$

It is quite obvious that these resulting vectors are not just multiples of each other, and so it is no longer true that pixels from the two regions differ only by a scalar and a constant vector. In order to find a useful function to determine the relative reflectance in a scene with multiple illumination colours, it helps to look at the flux space. We will now do this for two illumination colours.

Assume that the two colours are independent, for if one is simply a brighter version of the other, then the two act as a single colour for our purposes. Three channels will be used, but having more would not alter the principle. Consider a neutral surface under the two illumination colours, one at a time. Each source colour will produce a vector in flux space, as in the one source case. Any point on the vector is possible, due to changes in the local illumination brightness, up to the top value which is the full available brightness. Now when both are present at the same time, the resulting flux is the sum of these two values. That is, the sum of any point on one vector with any point on the other. From vector geometry it is well known that these points describe a diamond or

parallelogram, as shown in figure 3.4. The lower tip is at the origin, and the two adjacent sides are the single- source vectors. The far tip is the vector sum of the two single sources. This figure is important in later discussion, and will be referred to as the flux diamond.
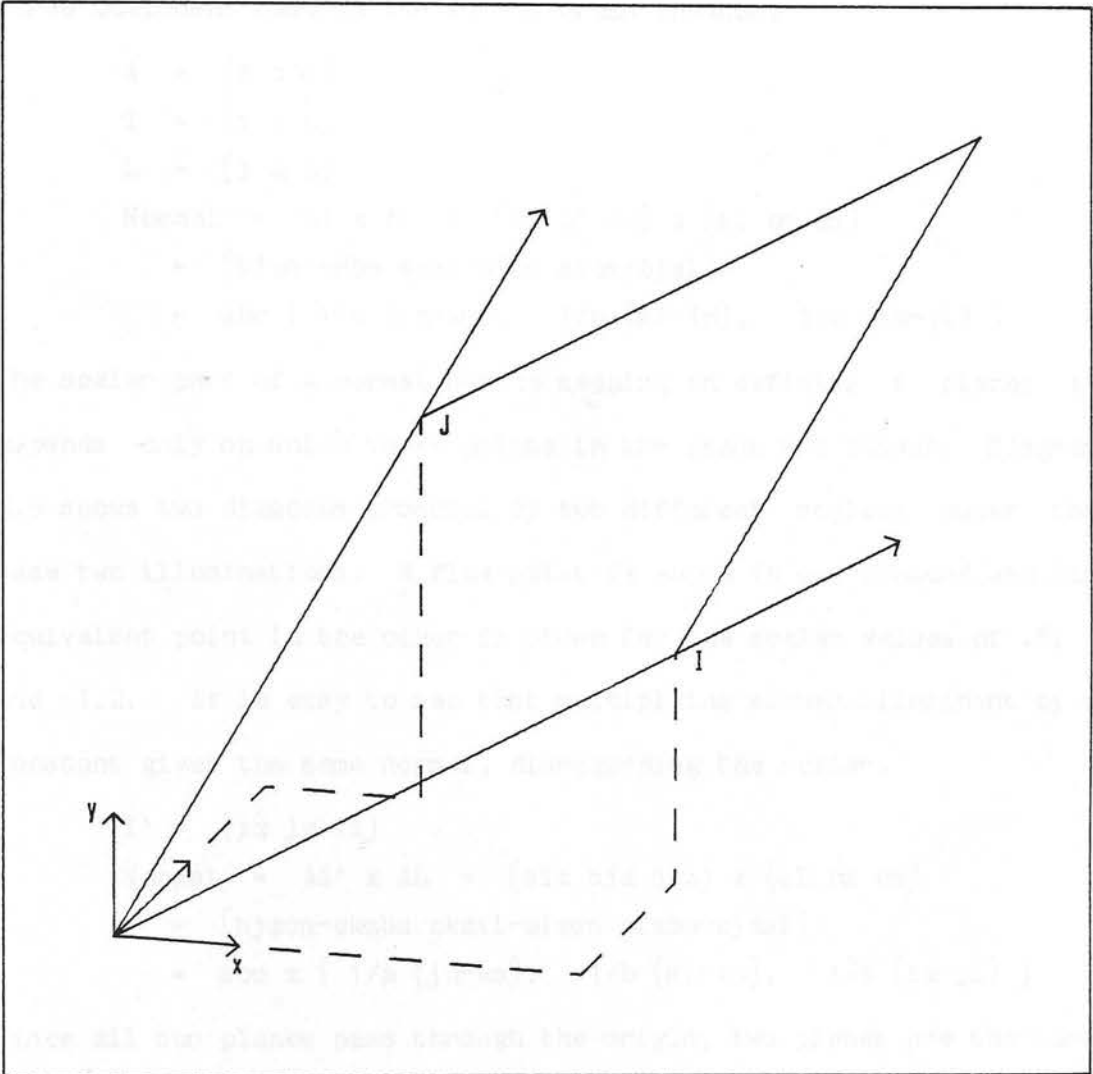
3.4 A Flux Diamond

So we now have two illumination colours on one region generating a flux diamond. When a new region of different reflectance is under the same sources, it will generate a different diamond. In fact, the relationship between these diamonds is directly related to their

relative reflectance. More generally, the ratio in the angle of the planes of the diamonds is a function of their surface reflectance ratios. The two vectors AI and AL, where A is the reflectance and I and L are the illuminations, together with the origin define a plane. The angle of a plane can be specified by its normal, which in the three component case is the vector cross product.

$$A = (a \; b \; c)$$
$$I = (i \; j \; k)$$
$$L = (l \; m \; n)$$
$$Normal = AI \times AL = (ai \; bj \; ck) \times (al \; bm \; cn)$$
$$= (bjcn-ckbm \quad ckal-aicn \quad aibm-bjal)$$
$$= abc \; (\; 1/a \; (jn-km), \quad 1/b \; (kl-in), \quad 1/c \; (im-jl) \;)$$

The scalar part of a normal has no meaning in defining a plane; it depends only on which three points in the plane are chosen. Diagram 3.5 shows two diamonds produced by two different regions under the same two illuminations. A flux point is shown in one diamond and its equivalent point in the other is shown for the scalar values of .5, 1 and 1.2. It is easy to see that multiplying either illuminant by a constant gives the same normal, disregarding the scalar.

$$I' = (iz \; jz \; kz)$$
$$Normal = AI' \times AL = (aiz \; bjz \; ckz) \times (al \; bm \; cn)$$
$$= (bjzcn-ckzbm \quad ckzal-aizcn \quad aizbm-bjzal)$$
$$= abc \; z \; (\; 1/a \; (jn-km), \quad 1/b \; (kl-in), \quad 1/c \; (im-jl) \;)$$

Since all our planes pass through the origin, two planes are the same exactly when their normals differ only by a scalar.
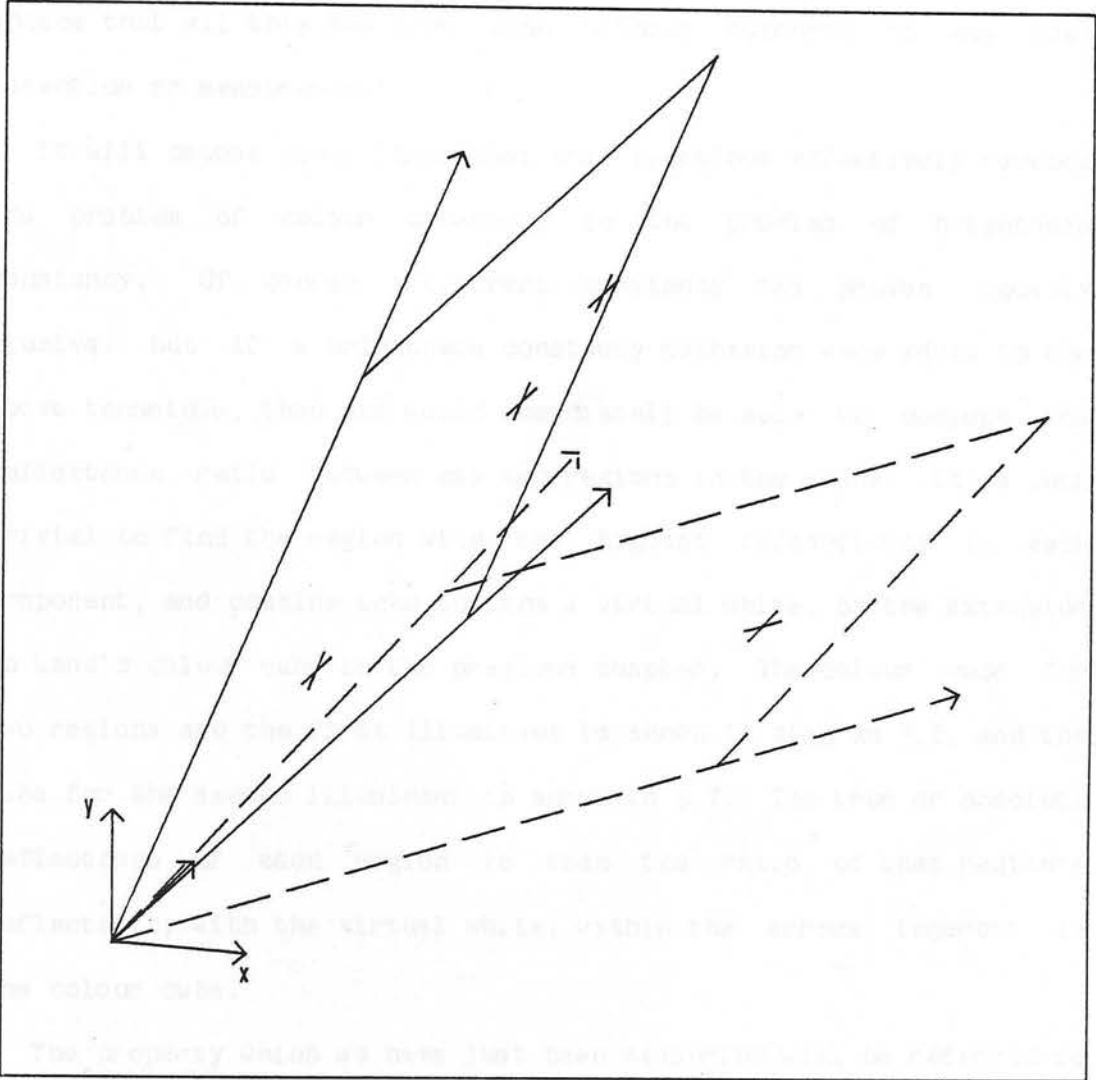
In order to compute the shift from one plane to another, take the component-wise ratio of their normals. We do this for the above region A and another D, under the same illumination colours, although

as just seen the illumination intensities may vary.

$$D = (d\ e\ f)$$

$$AI \times AL\ /\ DI \times DL$$

$$= abc\ (\ 1/a\ (jn-km),\quad 1/b\ (kl-in),\quad 1/c\ (im-jl)\ )$$

$$/\ def\ (\ 1/d\ (jn-km),\quad 1/e\ (kl-in),\quad 1/f\ (im-jl)\ )$$

$$= abc/def\ (d/a\ \ e/b\ \ f/c)$$

3.5 Fluxes Which Differ by Only a Scalar

So the vector ratio of the normals to two planes in flux space is the component-wise inverse of the colour ratio between their two reflectances, up to a scalar. This is a very useful result. If we take any two flux points from region A which are not co-linear with

the origin then we can compute the normal to the plane. Having computed the normal for D, we take the ratio of N(A) / N(D), and this is the ratio D / A, except for a brightness value.

$$N(A) / N(D) = AI \times AL / DI \times DL$$
$$= abc/def \quad (d/a \quad e/b \quad c/f)$$
$$= s (D / A) \quad , \text{ for some scalar s.}$$

Notice that all this has been done without recourse to any edge detection or measurement!

It will become clear later that this technique effectively reduces the problem of colour constancy to the problem of brightness constancy. Of course brightness constancy has proven equally elusive, but if a brightness constancy mechanism were added to the above technique, then one would immediately be able to compute the reflectance ratio between any two regions in the scene. It is then trivial to find the region with the highest reflectivity in each component, and combine them to form a virtual white, by the extension to Land's colour cube in the previous chapter. The colour cube for two regions and the first illuminant is shown in diagram 3.6, and the cube for the second illuminant is shown in 3.7. The true or absolute reflectance of each region is then the ratio of that region's reflectivity with the virtual white, within the errors inherent in the colour cube.

The property which we have just been measuring will be referred to as the transform. This refers both to the vector that is computed and the underlying property which is being studied, which is the ratio of the reflectances of two regions. Sample transformations are shown in diagram 3.8, where x is transformed by 2:1 while y and z are constant, and in diagram 3.9, where x is transformed by 2:1 and z by 1:2 while y is constant. For a moment imagine a scene where some

- 91 -

point in every region got the maximum amount of light from each
source. Then every diamond in flux space would have a cell at its
tip, and the ratios of these tips would be the true reflectance
ratios, including brightness! Furthermore, the component-wise
maximum of these tips gives the virtual white for the scene, making
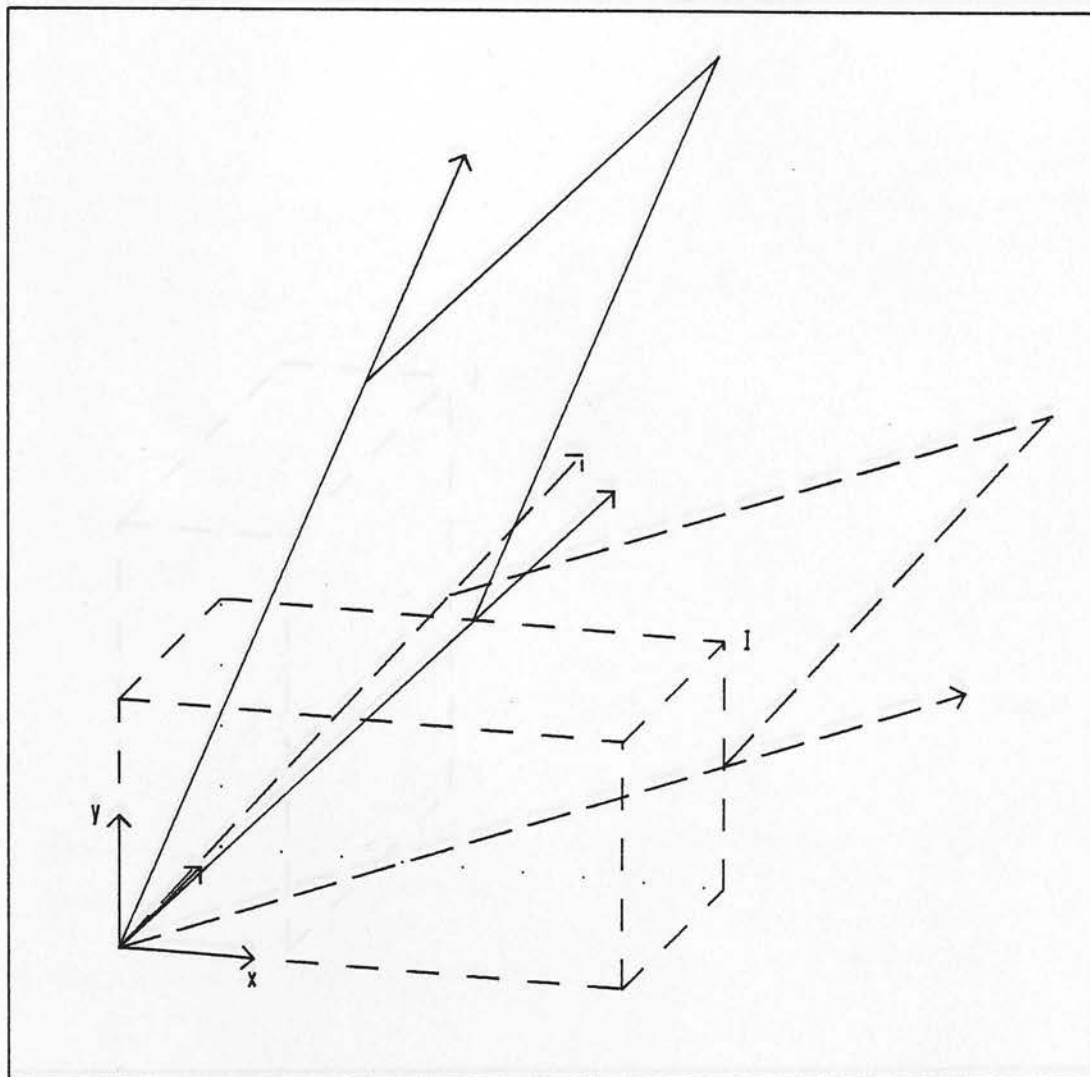
3.6 Colour Cube for One Illuminant

the flux space behave similar to a colour cube. Unfortunately, it is
not usual for every region in a scene to have some pixel in full
illumination, and so a lot of our subsequent work will be towards
compensating for the fact that every point in the diamond is not

- 92 -

immediately apparent; the maximum points in a region's flux plane
need not be due to the maximum illumination in the scene, although
they will be the maximum illumination in the region.

Knowing that a virtual white always exists, it will be useful to
think of every other region as a diminishing transformation from the
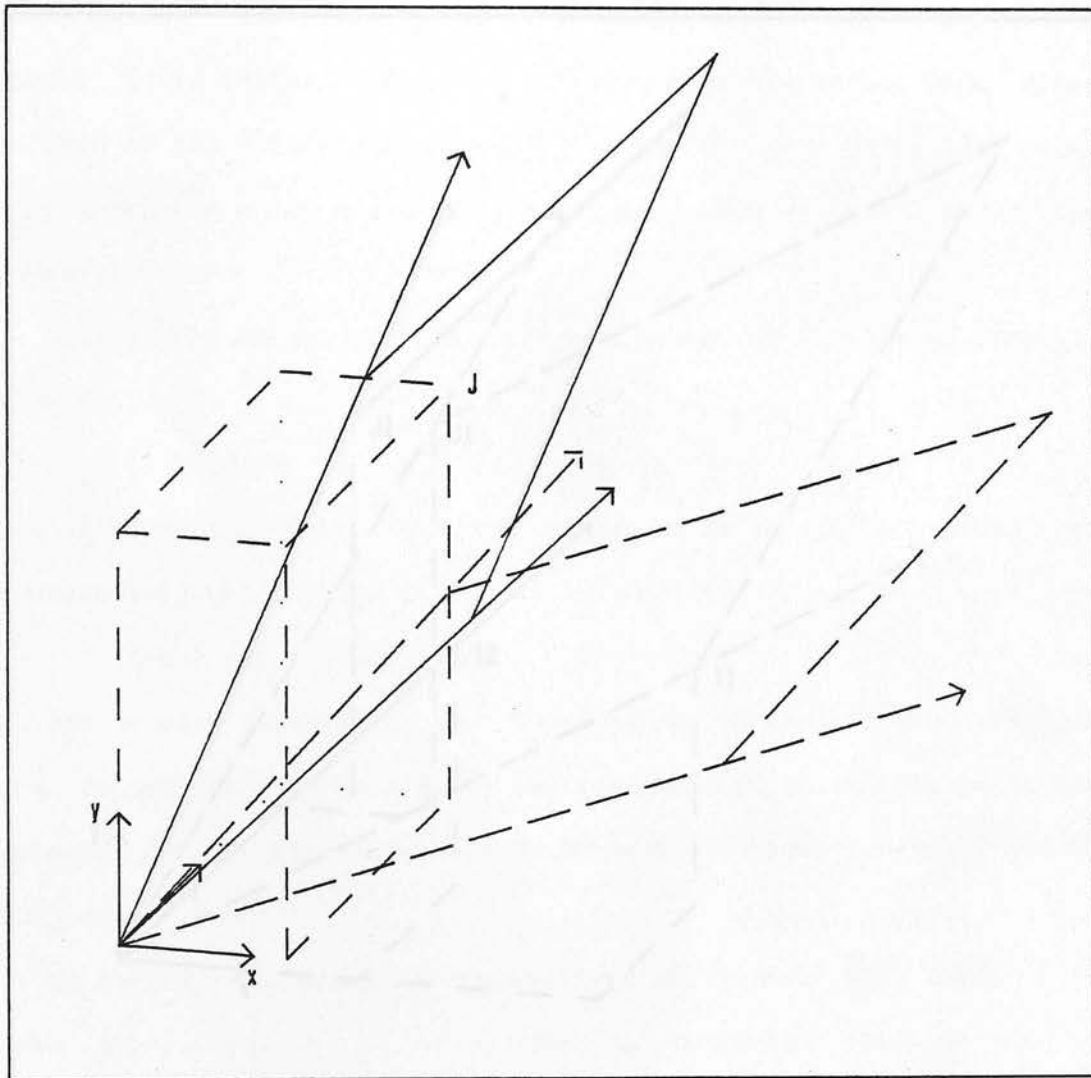
3.7 Colour Cube for the Other Illuminant

virtual white plane; that is, every component of a transform from
virtual white is less than or equal to one.

An interesting property of worlds which have multiple illumination
colours is that the structures they form in flux space, such as the

- 93 -

diamond and the plane it lies in, divide into equivalence classes. In the single illuminant case which began the chapter, each region produced a line segment in flux space which had one end on the origin. The transforms were the differences between the lines' angles. Given an illumination source, and so the angle of the
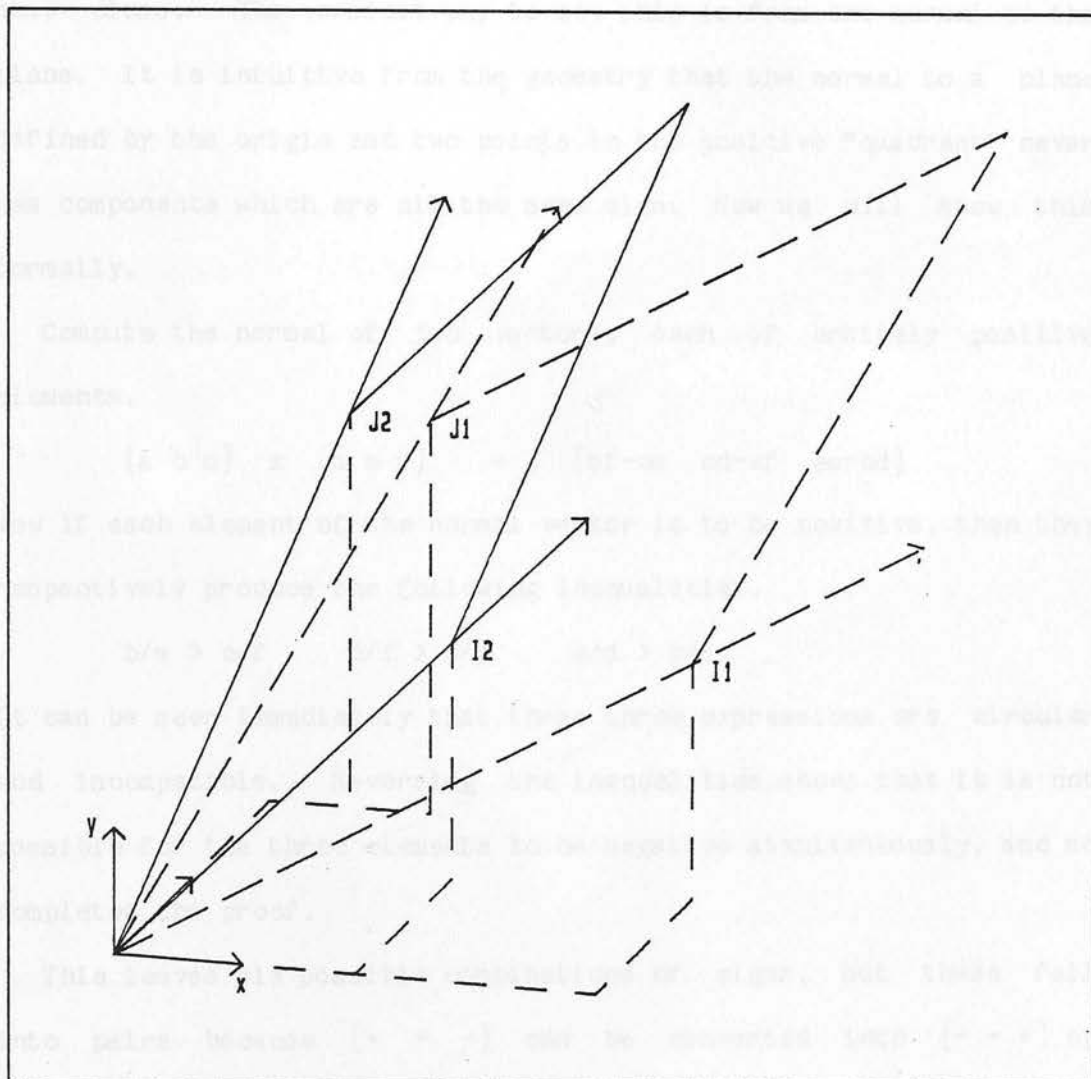
3.8 Transform with an x ratio of 2:1

"white" segment in flux space, it is possible to create a reflectance which will give any desired resulting angle. Since in this case every flux structure is obtainable within a single scene, there is only one equivalence class of these structures.

However, in the case of two illumination colours, and thus planar flux structures, the full set of possible structures is not obtainable under a single pair of source colours. Under three-receptor viewing, there are three equivalence classes of flux structures, and in any one scene all the structures will be from the same class. The easiest way to see this is from the normal to the plane. It is intuitive from the geometry that the normal to a plane defined by the origin and two points in the positive "quadrant" never has components which are all the same sign. Now we will show this formally.

Compute the normal of two vectors, each of entirely positive elements.

$$(a\ b\ c)\ x\ (d\ e\ f)\ =\ (bf-ce\ \ cd-af\ \ ae-bd)$$

Now if each element of the normal vector is to be positive, then they respectively produce the following inequalities.

$$b/e > c/f \qquad c/f > a/d \qquad a/d > b/e$$

It can be seen immediately that these three expressions are circular and incompatible. Reversing the inequalities shows that it is not possible for the three elements to be negative simultaneously, and so completes the proof.

This leaves six possible combinations of signs, but these fall into pairs because $(+\ +\ -)$ can be converted into $(-\ -\ +)$ by multiplying by a scalar of negative one, and scalars do not affect the normal, which is a line rather than a vector. The change of sign is equivalent to taking the cross product of the vectors in the reversed order. So the three possible normals are:

$$(+\ +\ -) \qquad (+\ -\ +) \qquad (-\ +\ +)$$

Finally, every element of a transform is positive, since it is the

ratio of two positive reflectances. Component-wise multiplying by a positive vector will not change the sign of the normal. This shows that every flux plane must have a normal of one of these types, and any other flux plane caused by a different reflectance must be of the same type. In essence, the relationship between the colours, but not
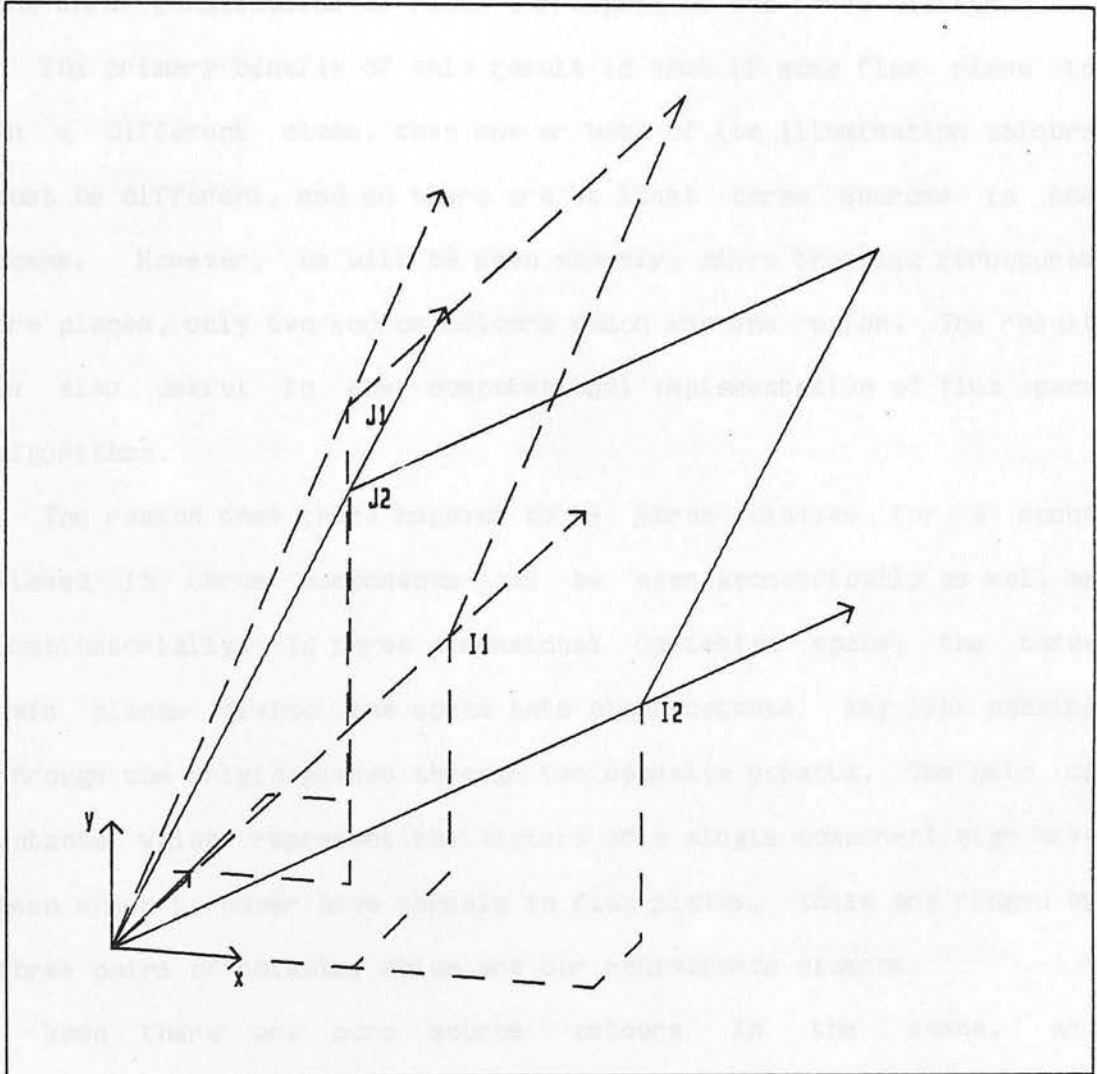
3.9 Transform of x 2:1, y 1:1, and z 1:2

brightnesses, of the two sources determines the flux space equivalence class of all the regions illuminated by those sources. The class can be determined immediately from the two colours; if the source colours are (a b c) and (d e f), then take the ratios a/d, b/e

and c/f. Whichever ratio is between the values of the other two is the position of the opposite sign. Notice that this also holds if the ratios are inverted. A geometric intuition on this topic is that any plane in the positive octant which passes through the origin will intersect exactly two of the three faces which border that octant; the three permutations of faces correspond to the three classes.

The primary benefit of this result is that if some flux plane is in a different class, then one or both of its illumination colours must be different, and so there are at least three sources in the scene. However, as will be seen shortly, since the flux structures are planes, only two source colours reach any one region. The result is also useful in the computational implementation of flux space algorithms.

The reason that there happens to be three classes for a scene viewed in three components can be seen geometrically as well as combinatorially. In three dimensional Cartesian space, the three axis planes divide the space into eight octants. Any line passing through the origin passes through two opposite octants. The pair of octants which represent the vectors of a single component sign have been shown to never have normals to flux planes. These are ringed by three pairs of octants, which are our equivalence classes.

When there are more source colours in the scene, and correspondingly more components in the imaging system, the set of flux space structures is divided into even more classes. In three source colours and four components, there are seven classes:

$$(+ + + -) \quad (+ + - +) \quad (+ -_{\cdot} + +) \quad (- + + +)$$
$$(+ + - -) \quad (+ - + -) \quad (- + + -)$$

In four sources and five components there are fifteen classes of flux planes, given two colours of illumination. The general formula for s

sources and s+1 components is two to the power s, less one. This is
because the s+1 components divide the space into 2**(s+1) partitions,
analogous to the octants, but these are grouped into pairs, and the
pair of all-positive and all-negative is excluded.

The next observation about planes in flux space has much wider
applications.   So far we have been using the uniqueness-to-a-scalar
-within
property of the flux planes. This is the bijection between planes
and colours of regions, which groups all regions of the same
reflectance colour together irrespective of brightness. Now we will
begin to use properties of the flux diamond which is in each plane.
Recall that the diamond begins at the origin, is bounded by the two
vectors from the two light sources, and ends at the sum of the two
sources. For the moment we will concentrate on the two lower
vectors, and call them A and B. These are the result of illuminating
the region with only one light or the other.

One of the most important properties of these vectors is that they
go to a unique vector under a given transformation of one plane into
another. This is trivial to show. If $(a\ b\ c)$ is a point on the
vector A, then all the points are on the line $n(a\ b\ c)$, for positive
n. The change of plane is by some transform $(x\ y\ z)$, which is
measurable from the ratio of normals. However, any scalar multiple
of this transform would also achieve the change of plane, so

$$T\ =\ m(x\ y\ z)$$

The transform acting on the line is the component-wise product

$$T(A)\ =\ m(x\ y\ z)\ *\ n(a\ b\ c)$$
$$=\ mn(xa\ yb\ zc)$$

and as the scalars combine, this is a unique line in the new plane.

The usefulness of this observation becomes apparent when we
realise that in realistic scenes the diamond will not be distinct on

- 98 -

each plane. If a region has some pixels which are in shadow, and thus receiving light from source A but not B, then the corresponding flux points will be on the A vector. Similarly, a shadow over the A source will give some points on the B vector. It is unlikely that each region will have both these two shadows within its bounds. It is much more likely that there will be at least one instance of the two shadows somewhere in the scene, probably in different regions. It is also likely that each region will have a non-constant flux; that is, some point in the region will get a little more or less light from some source than another point in the region. Such a flux change creates two points which are not co-linear with the origin, and so define a flux plane. With each flux plane, find the right-most and left-most flux points. These are respectively the candidates for the vectors A and B. The transform to the next plane is easily found from the normals of the two planes. Then these candidates for the outside vectors can be transformed onto the next plane. Which ever is outermost becomes the new candidate. By repeating this procedure through all the planes, the leftmost and rightmost of all the flux points will be found. Given that each of the sources is somewhere in shadow, the shadow will be found. These two shadow values can then be directly propagated back onto every region, giving the bottom "V" of the diamond.

Now we know the direction, but not yet length, of the A and B vectors on each region. Within a given region this allows the flux of each pixel to be divided into its components of flux due to source A and flux due to source B. Algebraically, the light sources are known to be independent vectors, or they would be co-linear with the origin and behave as the same colour. So another vector in the space

- 99 -

has a unique representation as the weighted sum of the two vectors. Begin by choosing a point on each of the A and B vectors.

$$A = (a\ b\ c)$$

We will determine how much of these two vectors is in a given flux point F.

$$F = (g\ h\ i)$$

Later it will become apparent how to most effectively choose the values of A and B, but any point on the vector will work. We are looking for values of n and m which solve the vector equation

$$n\ A + m\ B = F$$

which is equivalent to the redundant simultaneous equations

$$na + md = g$$
$$nb + me = h$$
$$nc + mf = i$$

In fact it is only necessary to have two colour components to solve this system, finding for instance

$$n = (ge-hd)\ /\ (ae-bd)$$
$$m = (gb-ha)\ /\ (bd-ae)$$

So this gives the amounts of A and B, respectively in any measured pixel. That is to say that if all the sources of colour B were turned off, the flux at this pixel would be

$$F(A) = n\ (a\ b\ c) = [(ge-hd)/(ae-bd)]\ (a\ b\ c)$$

and similarly the amount of flux due to only the sources of colour B, without those of colour A, is

$$F(B) = m\ (d\ e\ f) = [(gb-ha)/(bd-ae)]\ (d\ e\ f)$$

We will now show that these results are independent of the choice of A and B. It is sufficient to show that the same result is obtained from an arbitrary A. Any choice of A on the vector will be a scalar multiple of this A, so

$$A' = x A = (xa\ xb\ xc)$$

Computing as before gives

$$n' = (ge-hd) / (xae-xbd) = 1/x\ n$$
$$m' = (gxb-hxa) / (xbd-xae) = x/x\ m = m$$

So finally the x terms cancel in the flux formula.

$$F(A)' = n' A' = 1/x\ n\ x\ (a\ b\ c) = n A = F(A)$$

There is a natural geometric interpretation for this method of separating the effects of the fluxes from the two source colours. In the flux plane the two source vectors, A and B, are independent and form the axes of a skewed Cartesian system. Within the flux diamond, all lines parallel to the B vector represent the fluxes of equal amount of A illumination, and any line parallel to the A vector covers points of a constant amount of incident light from source colour B.

Notice at this point that this provides a useful mechanism for applying the tools of photometric stereo [Woodham, 1978]. In monochrome photometric stereo, scene properties are found by looking at a scene with one illumination source on, and then viewing the identical scene under a second source which has a different position, while the first source is off. The derived information is based on the different angles to the sources and the resulting changes in patterns of illumination and shadow. Obviously the monochrome version requires that the scene be motionless between the two images, and that the two lights flash on and off, with the attendant strain on both mechanisms and human observers. A polychrome version would assign a different colour to each source, and leave each source illuminated continuously. Using one of the imaging systems which takes polychrome pictures, the scene could be in motion, and the colour flux processing just described would separate the two sources.

There would be no distracting flickering of lights on and off.

Although our mechanism has provided a way of separating the effects of two different colours of light source, we do not yet have a brightness constancy mechanism. In particular, if a cell in one region has a flux due to source A of n A, and a cell in a different region has a flux due to source A of p A', where A' is the value of A after the transformation from the reflectance vector of the first region to that of the second, these two values are not directly comparable. Since we have only established the colour ratio between regions, and not the brightness ratio, there is an unknown scalar in the transformation; we do not know that the second cell receives p/n times as much light of colour A than does the $first$ cell.

However, what we do know is the ratio of light due to each source. That is that $F(A)/F(B)$ is independent of the brightness scalar, and so is directly comparable between any two regions. We will show this now. Given the transformation T, from one region's diamond to the other

$$T(V) = x(s\ t\ u)\ V$$

where the values of the elements of the vector $(s\ t\ u)$ are known, but the scalar x is not, apply it to the flux ratio. This ratio is n/m, the ratio of our two scalars from the first proof. Remember that each of these scalars is independent of the choice of point on A and B, and independent of whatever units are used to measure A and B, provided that both vectors are measured in the same units.

$$
\begin{aligned}
T(n/m) &= T(nA)\ /\ T(mB) \\
&= x(s\ t\ u)\ nA\ /\ x(s\ t\ u)\ mB \\
&= n\ (s\ t\ u)\ A\ /\ m\ (s\ t\ u)\ B \\
&= n/m\ T(A)\ /\ T(B)
\end{aligned}
$$

For the present purposes it is important that the scalar component of the transformation cancels out, and leaves the ratio unchanged on the

next region. In fact, the entire transformation can cancel out, but this property is only the constancy under reflectance transformations of certain ratios within the flux diamond.

The geometric interpretation of the constant ratio n/m is that any angle within the flux diamond is *proportional* when the diamond is transformed to a new reflectance region. However, we will soon see that the angle of the diamond is altered in a predictable fashion by the transformation. Therefore, for an angle within the diamond to be invariant, it must be expressed as the ratio of the angle to vector A with the angle to vector B.

In the next two chapters we shall use the ability to separate the effects of the two source colours, firstly in a world of distant point sources to derive information about the geometry of the scene, and secondly as input to polychrome edge-measuring algorithms.

Having now established the A, B coordinate frame in each flux plane, it is possible to find the maximum amount of flux from A in the region, and similarly the maximum amount of B. These flux values provide a local minimum for the upper edges of the flux diamond. The actual diamond must include this one.

This local diamond can be the basis of a naive colour cube. The "naive colour cube" does not attempt to compensate for the fact that some regions do not receive full illumination at any point. It initially treats each region as if it were in full light, by assuming that the brightest fluxes from each region are directly comparable. The naive colour cube can be obtained directly from the flux space, as follows. For each plane in flux space, take the two outside corners of the local diamond. These correspond to the amount of flux due to A and the amount due to B, respectively. Find the component-

- 103 -

wise maximum of the A values, and call it the naive virtual white for A, and compute a similar value for B. This is the same as finding the corner of the minimum cube which contains all the A's, or respectively B's. If, for each component, and each source, some region which is brightest in that component gets full light from the source then both virtual white values will be correct. In a rich scene, with many different colour regions, each getting a range of brightnesses, the naive colour cube heuristic should provide a reasonable approximation. However, having approximated virtual white is not the same as having determined the brightnesses of every region. Within this naive colour cube, the maximum value for the flux provides the lower bound for the region brightness; this is equivalent to the assumption that the distance between the region's maximum flux and the surface of the colour cube is entirely due to the weak reflectivity. The opposite assumption is that only dim light is arriving at the region, which is maximally reflective; this provides the upper bound for region brightness, which is the intercept of the vector with the colour cube. The results of the naive colour cube may be combined with those of the "colour cube core" heuristic proposed in chapter two to refine the results. Still, the techniques are only suitable for quick approximations in restricted scene domains. Our main interest in them is to provide a groundwork on to which brightness information can be added to compute precise reflectance values.

We will now look briefly at different dimensionalities of scenes and imaging systems. Most of this chapter has used the example of two source colours under three components of receptor. The techniques may, however, be generalised to any case of s source

colours and s+1 components. It is instructive to see what is lost and what is retained when *fewer* components are available. Consider the case of two channels and two sources. The flux structures will still be diamonds, but they will be in the same plane, and often overlap. This immediately disrupts the region finding mechanism, since a border point will share the same plane with both adjacent regions. This can only be solved in a general scene by making additional assumptions about the type of flux changes which are likely to occur between adjacent cells of the same region, and so comes close to traditional monochrome edge detection. Also lost when the flux structures become co-planar is the automatic alignment of the diamonds; the direct method of determining the positions of the pure A and pure B vectors in a given region would no longer be available. Still, for regions with known diamonds, such as those where a shadow is found, the transformation between regions is directly computable from the transform of the diamonds. In addition, these regions would permit the construction of a local diamond, and the subsequent processing which follows from it. Notice that the width of a diamond is pre-determined by its angle with respect to the axes of flux space. In particular, any diamond has its widest angle when the diamond is bisected by the 45 degree line between the axes. It diminishes to nothing as it approaches either axis. This is independent of the colours of the illumination, or the width of the diamond at the point of virtual white. Probably there are various tricks which can be used to make use of the partial information contained in each ambiguous flux structure. For example, a few regions with shadows would establish a copy of a correct local diamond. It is then known what the diamond will look like at any

angle in the plane. This defines the entire class of possible diamonds in this scene; a given blob of fluxes will only fit within a limited range of these diamonds, and this range is much smaller than would be expected without knowledge of the behaviour of the diamonds under transformation. However, these and similar tricks can at best only produce the information which is available easily if one additional receptor is employed.

Recall that when separating the effects of the two sources, a system of simultaneous equations was solved. This system underlies most of our colour constancy results, and so requires that there be at least as many components as there are source colours. Virtually no results are obtainable with fewer receptor channels, because every region can easily span the entire flux space.

At the opposite extreme, what will happen when the number of components is two or more greater than the number of source colours? Clearly this will not be a hindrance, since it is always possible to ignore the values of the excess receptors, and so reduce the flux space to an already solved problem. There is certainly an advantage in having an extra channel, in that it confirms the model of the lighting. If there was another unexpected source colour, then the extra receptor would be able to detect it, and if necessary process the scene as having one more illuminant than expected. However, the handling of cases larger than three dimensions is not as straight forward computationally. For example, consider two sources under four dimensions. Clearly the flux structure is still a diamond. However, it is not as easy to describe the plane which the diamond occupies. In the three dimensional case there is the normal to the plane; it describes the plane precisely, there being one normal for

each plane. In four dimensions, the plane does not have a unique normal; the field of possible normals itself occupies a plane.

There are a variety of ways of solving this problem of unique representation. One classic method is to use Plucker co-ordinates in projective space. It is essentially a way of removing a dimension from a space, in this case the flux space, so that in the new space each structure has one less dimension. Details of Plucker co-ordinates can be found in Todd [1947]. However, these general tools are unwieldy in this limited application; our present domain has the simplifying traits that all the planes are in the positive partition of the space, all of them pass through the origin, and importantly, none is parallel to any axis plane. In fact, any ad hoc method for determining what vector will perform the transformation between two given planes can be made to serve. An obvious candidate is the ratio of intercept points with appropriate lines in the axis planes. This finally falls under the heading of implementation details.

For completeness, we have a quick look at the geometry of three dimensional structures in four dimensions of receptor. There will be three source vectors, and so the flux structure will be a parallelepiped. Each of the three sides which border the origin will be the flux diamond produced by two of the sources in the absence of the third. This structure is embedded in a hyperplane which has a unique normal, and so is easily measurable and manipulatable. Once again it is only necessary that, for each source colour, somewhere in the scene there is a pixel which is in shadow from that source for there to be an easy definition of the flux structure in every hyperplane. Similarly, all the computations of this chapter follow easily. The following two chapters rely upon playing off the

brightnesses of the different source colours to extract further information. For example, if when comparing two regions, the A drops by 10%, and the B drops by 20%, then surely the extra amount missing from the B is entirely due to reduced lighting; that is, the first 10% can be explained by either lower reflectivity or less light in both A and B, but if the second 10% were also attributed to low reflectivity, the region must receive 110% of the available light from source A, which is clearly impossible. Clearly, when a third light source is added to these types of computations, it can only enhance the final results.

We will return now to the standard case of two sources and three dimensions, and show how to handle a few apparently difficult ambiguities.

Even under two sources, it is possible for the flux in a region to be so impoverished that it defines only a line rather than a plane. This can happen in three ways. The most common is that the entire region is in shadow relative to one of the sources. Typically the set of flux points will define a line of some length, rather than just a single point. Also it is likely that the adjacent pixels of some neighbouring region are also in the shadow. The second case is a very small flat region, where all the pixels give the same flux, and so the flux set is a single point. The third case is where the changes in flux of A and B are perfectly matched in co-incidental alignment so that they cancel each other out to vary the flux only by amounts of grey. This last case is only likely in contrived artificial scenes. The first two cases are easily distinguished, both by their shape and by whether their neighbours are in shadow. Now knowing that a line is due to a shadow means that it must align

under transformation with one of the two edges of the flux diamond. This limits it to two transformations, one of which will yield very unreasonable values in later brightness and colour cube computations. Further, it will almost certainly align with the side of the diamond that its neighbours which are in shadow will occupy. Therefore, a degenerate flux structure which is due to a shadow is easy to recognise and process. However, the case of a region with a constant flux from every pixel is more difficult. Generally it seems better to ignore that region until all the other regions have been processed to reveal reflectance colour and lighting. Several heuristics then become available, for instance, if all of the border pixels are in a narrow range of illumination, then the region is probably under the same illumination. If such troublesome regions are frequent, it is a sign that the imaging receptors are not sensitive enough; better equipment would detect the flux changes which occur in almost every region that is larger than one pixel.

Another possible ambiguity occurs in the region finding stage. Typically all of the pixels within a region will be on the same flux plane, and those in an adjacent region will be in a different flux plane, and so the location of the edge is immediately found. We are for the moment ignoring mixed pixels and the problems of specific imaging systems. However, even in an ideal scene, there is still the possibility of an ambiguous cell at the border between regions. This is because the flux planes will usually intersect. Should a flux value happen to lie on the intersecting line between the two flux planes, and also in the scene on the border between the two regions that the planes represent, then it is initially ambiguous. There are two obvious way to resolve its location. Firstly, mark it as

unspecified, and return to it when the locations of the flux diamonds in each plane have been determined. It is likely that it will fall within the diamond of only one plane. Secondly, compare it with the adjacent cells in each region for continuity. Having located the diamond in each plane, it is possible to compute the amount of flux from A and B which would be attributed to it if it were in one plane or the other. These amounts should be reasonable with respect to its neighbours in the scene region. Should it pass both these tests then it is truly ambiguous; for the purposes of colour processing it could be considered to be in either region, and no other known test could disambiguate it.

We will now turn to examine an actual implementation of flux space methods. It was built to demonstrate their viability, and also for the practical lessons that occur when sufficient detail must be specified to make a theory work. The programs were limited to the cases of one and two illuminant source colours only although no difficulty was seen in extending them other than the bulk of data and the increased likelihood of obscure mistakes in the programs. The programs primarily used an input of four colour separations for convenience in determining the dimension of the illumination. Several of them were run on three colour data in the early stages without a change in result.

In the earliest stages of the present project, before the development of the present mechanism, considerable processing was done with digitised images of real scenes, taken through a monochrome camera and filters. It was eventually concluded that that data was unsuitable for the present purposes. One hundred grey levels could be obtained only with difficulty, and noise was in excess of six

levels. In addition, the gain with brightness varied unpredictably in a typical ten minute session to photograph all the colour planes for one scene. The spectral response in the blue-geen half of the spectrum was very poor, and never more than twenty usable levels were obtained. This collection of difficulties gave insights into the problems of practical vision systems, and served the very important function of turning the author away from ad hoc techniques towards a theoretically sound solution. As a result, all of the programs described in this and subsequent chapters were applied to simulated data, created by an entirely separate set of programs from a description of the objects of the scene, the light sources and the camera position. An immediate advantage of this is that any failure in result could not be blamed on bad data, and so the cause was found all the more quickly in the implementation itself. It was also possible to quickly create an image of a scene of any type necessary, and to add any desired amount and type of noise. Furthermore, the pixel maps of reflectance and illumination were computed separately before being combined, and copies were output for direct automatic comparison with algorithm output.

No analogy is intended between the details of this electronic mechanism for reflectance constancy and the biological mechanism. Any computer programmer knows that a given solution can be arrived at many different ways on the computer. It may be all the more different when computed by the complexities of the neurology of the retina and cortex. The programs built here were not designed for an end result, but rather were a workbench of tools that could be combined to test different hypotheses and methods. As a result of this modularity, the algorithm has an unnecessarily sequential

layering. An implementation built for use rather than experiment would benefit from increased interaction between the layers. Of course no claims for speed or efficiency are appropriate.

The first stage is determining the illumination dimension. For the experienced human observer looking at image data in a flux space representation, the dimension is immediately obvious, since the data clumps into either lines, planes or subspaces. This is surprisingly hard to do reliably for all cases by serial computer. It does not seem appropriate for the computer to plot the points into flux space and then attempt to find lines and planes in the image. Instead, small areas of the scene are studied to determine the local dimension. However, if the area includes a reflectance edge, the dimension increases, and a junction of reflectance edges can increase it further. Recall that at this stage the location of the edges is not known, and to assume them would be circular. This is an area where the method would be easier if it were less modular. The illumination dimension could almost certainly be computed simultaneously with the next step of edge and region finding.

Reflectance edges are not the only difficulty. If a region is in shadow then it has one less dimension. Also troublesome are flat areas and very slow changes of brightness. In a flat area the illumination strength and thus the flux may remain constant, and so remove all clues to illumination dimension. With a slow change of brightness in a digitised image, the value may change from one grey level to the next only at intervals of several pixels. To many tests these occasional changes appear to be weak shadow edges. It is interesting to note that flat surfaces are regarded as a virtue by most monochrome vision programs that pre-date the shape from shading

paradigm [Horn, 1975; Pentland, 1982b,c; Smith 1983].

Although subsequent processing was only built for one and two source colours, the first stage was required to recognise when the dimension was greater than two. This prevented the first stage from degenerating into a guess between one and two sources. In fact, most of the complexity arises in distinguishing two from more than two.

One solution to determining the dimension is to use a mask which blindly determines the local dimension at every image point. A histogram of these values, which are all small integers, is then taken. Since reflectance edges are rare in comparison with pixels, the dimension should be the most frequent number in the histogram. While this was confirmed experimentally, there are several reservations. This method fails on scenes with texture, lots of edge detail, copious shadows, or mostly planar surfaces. It is not uncommon to have two adjacent histogram values as possible candidates, and without a priori knowledge of the type of scene, it is not possible to prefer either one.

The alternative chosen is for the mask to do a certain amount of guessing as to region boundaries and the reasons for local changes in dimension. This was initially done by distinguishing the cross sections of the different edges, but later it was preferred to find a canonical set of basis vectors for the local pixels. Thus as the mask travelled across the scene, many changes simply added or removed a vector. A reflectance edge would add its vector only along its length, whereas a surface under gradual change would have the same basis across its area, with possible local deletions. This increased program intelligence was able to solve all but the most contrived scenes.

The next stage was to definitively find the regions and edges. Because of the way it was developed, two different programs existed for one and two dimensional illumination, depending upon the results of the first stage. The one dimensional case is easy, since adjacent points either are or are not on the same line.

With two source colours there are several problems, since the plane has more freedom to move about. Although in theory even a small amount of variation in the flux points is sufficient to define the angle of the plane, the possible error is large. The present program searched for a region which had a wide variation in flux, determined the plane which fitted that data, and then used the transformations of that plane which passed through the data of less variable regions to define the exact plane angle for those regions. A more refined method might compare all the regions and find the statistically best fit collectively.

In both dimensions at this stage the algorithm also looked for mis-fits. Data that produced negligible sized regions or planes for which no reflectance transformation was possible are signs of incorrect processing, such as an undetected edge. The latter could also be caused by a contrived scene in which the second source colour in one part of the scene was different from the second source colour in another part. Since the three sources were nowhere concurrent, the scene was processed as a two source scene, yet no transformation is possible for the second source. Although data checking such as this was built into every stage of the algorithm, the error was always processed manually so that its cause could be found and reduced. An applications program would process errors automatically when possible.

Having now the regions and the transformations between them, processing continues for the multiple illuminant case, and the third stage finds the actual source vectors relative to each region. That is, it finds the "V" at the bottom of the flux diamond. This is in fact straight forward, first finding the angle of each flux point on its plane, selecting the two extremes in each plane, and then finding the extremes among these by transforming the pair from each plane into a standard central plane. Error checking at this stage consists of checking that areas identified by shadows in the first stage all fall onto one of these lines. Typical misses are due to cumulative computational errors, and these can be substantial for narrow flux planes. Correction of errors beyond this tolerance band depends upon implementation, but should not be necessary after complete debugging.

The fourth stage is to find the upper edges of the diamond in each region, which again is easy to build. The "V" is transformed onto each plane. It presents essentially two linear equations in two variables, being the amount of each source present in each flux point. For all the points in the region, the largest solution value in each variable defines the height of that upper edge.

After this stage the processing becomes similar to previous colour cube computations, and is nearly the same for either illumination dimension. From the flux data are extracted the transforms, which tell the reflectance ratio between regions, and the height of the upper edges, which give the minimum incident brightness. A separate input gives the lightness ratio between some regions, and virtual white is computed from a linked network structure of regions. The virtual white is then passed back through the transformation, scaling each diamond to give region reflectance. The pixel brightness then

scales the region reflectance to compute incident pixel flux per source colour. Both reflectance and flux were then output and compared by a separate facility, at a rounding tolerance level, with the actual values.

The separation of the brightness stage allowed it to be initially put in by the artificial image program, thus simulating results from other visual processes such as perhaps spatial processing. As discussed in chapter five, it did not require that every edge have a lightness ratio, but only that lightness ratios chained together substantial groups of regions by some route. The modular structure meant that later, the edge classification measure of chapter five could input the reflectance ratio. Unreliable edges, as discussed there, would be reported simply by the absence of a ratio, thus preventing chaining across that edge.

So in this chapter we have provided a mechanism for separating the effects of different colours of illumination sources, which is useful for photometric stereo, and for several algorithms to be considered subsequently. Whereas previously the addition of more colours of source lights was thought to make the scene almost impossible to process, it is now seen that the effects of each light can be isolated, and so additional illumination colours will not reduce the amount of information which can be extracted from polychromatic scene processing, and will usually increase it.

We have provided the basic vector arithmetic of colour vectors, which will be used throughout the thesis. Region finding has been seen to be a natural consequence of seeing the flux space under an appropriate number of channels. This technique is useful on its own, but is also essential input for all subsequent stages. Within the

- 116 -

flux plane specified by a given region, the diamond produced by two illuminant colours has equivalent flux structures in scenes with more illumination colours. However, continuing with the two-colour example, the next chapter will show that the entire diamond can never be filled in a scene with a single distant point source for each illumination colour. When this fact is combined with the techniques we have just seen for separating the effects of sources A and B within the scene, we will have a way of determining the normals to the surfaces in the scene relative to the two sources. These should not be confused with the normals to structures in flux space. The relative surface reflectance information will also be used in chapter five under both the naive and enhanced colour cube constructions in order to compute the true white within the scene, and so the absolute reflectance of each region.

Relative Angle of Sources from Flux Space Envelopes

Determining Surface Normals from Colour

Colour Constancy under Three or More Sources

Controlled Illumination

In the preceding chapter we demonstrated a way of separating the effects of different coloured sources in a scene. Provided that the source colours are independent relative to the spectral responses of the channels of the imaging mechanism, and provided that there are more channels than there are source colours, then it is possible to produce a set of images which are the same as those that would have resulted from viewing the scene under each of the illumination colours in turn, in the absence of the others. This result did not assume any physical characteristics of the sources, other than independence of colour. A given source colour could be common to many different sources illuminating the scene, and they could be of any position or intensity. Linear sources, diffuse background illumination and focussed point sources all have the same effect with respect to this algorithm.

In the present chapter we will restrict ourselves to distant point sources in order to extract certain spatial characteristics of the scene entirely from the colour information. Recall that the changes in flux within a region are due to changes in the amount of incident light reaching it from one or more sources. There are three basic causes of this change in brightness. Firstly, as the region moves away from a source, it receives less light. Secondly, as the angle of the surface to the light turns away from perpendicular, the

surface receives less light. Finally, shadows reduce the amount of incident light, either cutting off a source entirely, or in the case of unfocussed shadows, reducing it across the width of the penumbra.

The reason for considering only distant point sources is that this reduces the causes of flux changes. When the source is distant, then the changes of brightness due to distance from source are insignificant, and so the first cause is eliminated. A point source is effectively focused, and so removes the possibility of gradual shadows, the penumbrae. There remain only abrupt shadows, where the region is not lit by a particular source, and changes due to the angle of surface with respect to the source. Further, a distant source means that the light is effectively arriving from a single direction at every point in the scene, and so removes parallax problems.

We begin with the assumption that each illumination colour comes from only one source, which is distant. By looking at the flux structures it is possible to determine the number of sources; if the structures are all planes, then there are two sources, if the structures are three dimensional, then there are three sources, and so on. We will now examine a way of determining the angle between the sources, relative to the scene. That is, if two lines are drawn from a pixel, one to each source, the angle between these lines can be closely approximated by studying the flux structures in the scene. To reduce confusion, this will be explained solely in terms of the two source case.

Recall the layout of the flux diamond. The two edges adjacent to the origin are the A and B vectors, and will also be called the lower edges. Each flux point on the A vector corresponds to that

- 119 -

proportion of light from source A, with no added light from source B. Now the two upper edges to the diamond express the opposite extreme. The one that goes from the end of the A vector to the apex of the diamond, which we shall call the "upper A edge", represents full flux from source A plus differing amounts of flux from source B. The apex corresponds to simultaneously maximum flux from both lights. We will now show that in the distant point source world, the entire diamond can never be filled in any one scene.

Consider first an extreme case. If both the light sources are on the same line from the region, then they will act together; as a surface turns away from one, it is also turning away from the other. If the surface is perpendicular to the source line, then the flux will fall at the apex of the flux diamond. If the surface is turned partially away, then the flux will fall somewhere on the line between the apex and the origin. When the normal is perpendicular to the sources or beyond, the flux is zero, and the flux point is at the origin. Notice that the only points in the flux diamond which can result from this physical layout are those along the main diagonal. It is not even possible for one light to be in shadow without the other being occluded as well. Notice too that it is possible to measure the cosine of the angle between surface and source as the distance along this main diagonal.

Now consider a more typical case, where the two sources arrive at the scene from two different angles. If a surface is perpendicular to source A, then its flux will lie somewhere on the upper A edge, depending upon its angle to B. Similarly, a surface point normal to source B will produce a flux on the upper B edge. However, it is not possible for the surface to be normal to both sources simultaneously.

- 120 -

This means that the apex of the diamond will never be represented in any flux diamond in this scene. It will be seen shortly that in this world of single distant point sources there can never exist a scene which fills every point of a flux diamond. In particular, areas on or near the upper edges of the
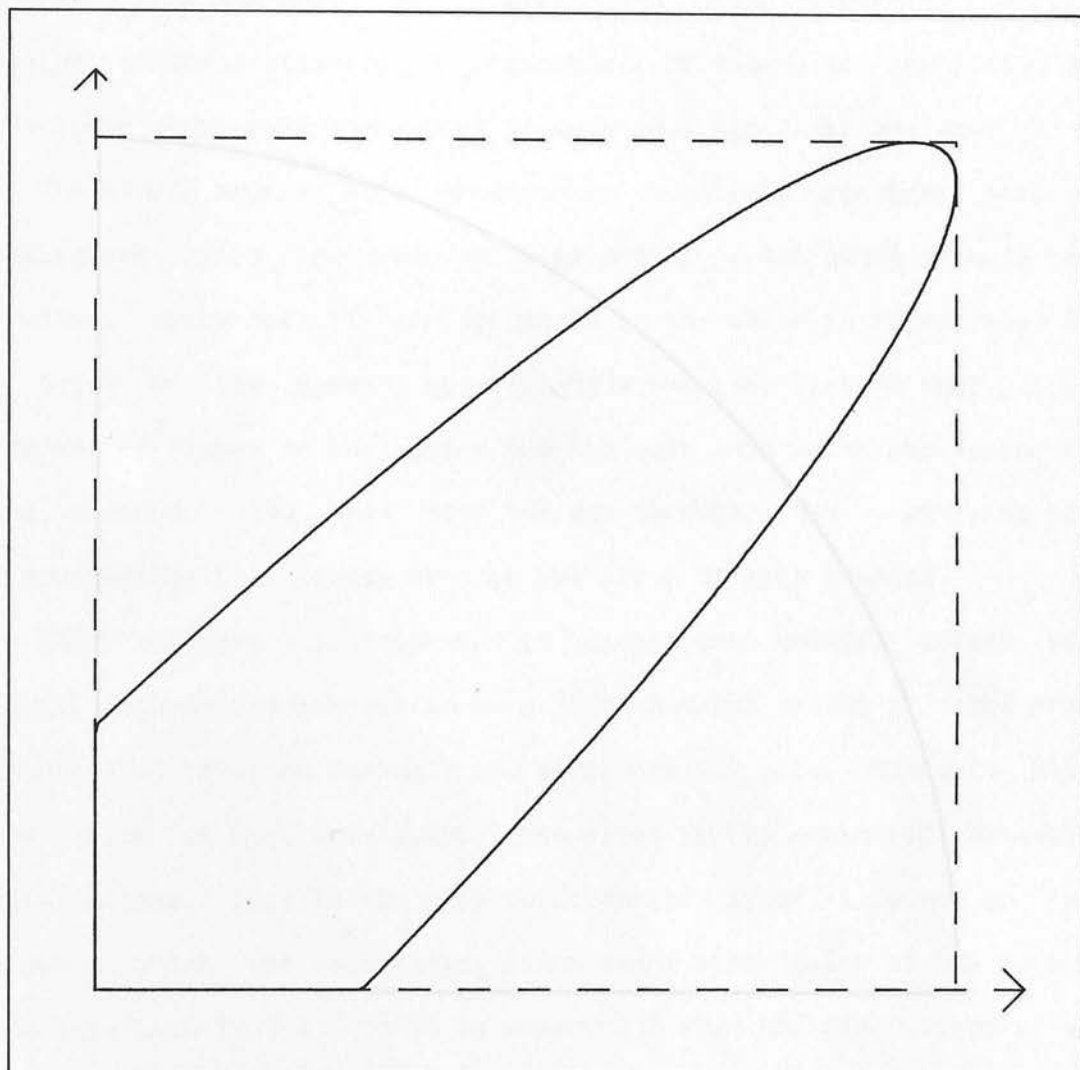
4.1 Envelope of angle .1π radians

diamond can never be produced. Which points are not producible is determined entirely by the angle between the two sources. The line of maximum obtainable flux points within the diamond will be called the envelope. Each source angle has a characteristic envelope, and

the envelope is common to all diamonds produced in the scene,
independent of reflectance transforms. Diagram 4.1 shows the
envelope characteristic of a source angle of .1 radians, or 18
degrees. In all of the envelope diagrams, the diamond has been
expanded to a right angle in order to better show detail. This is
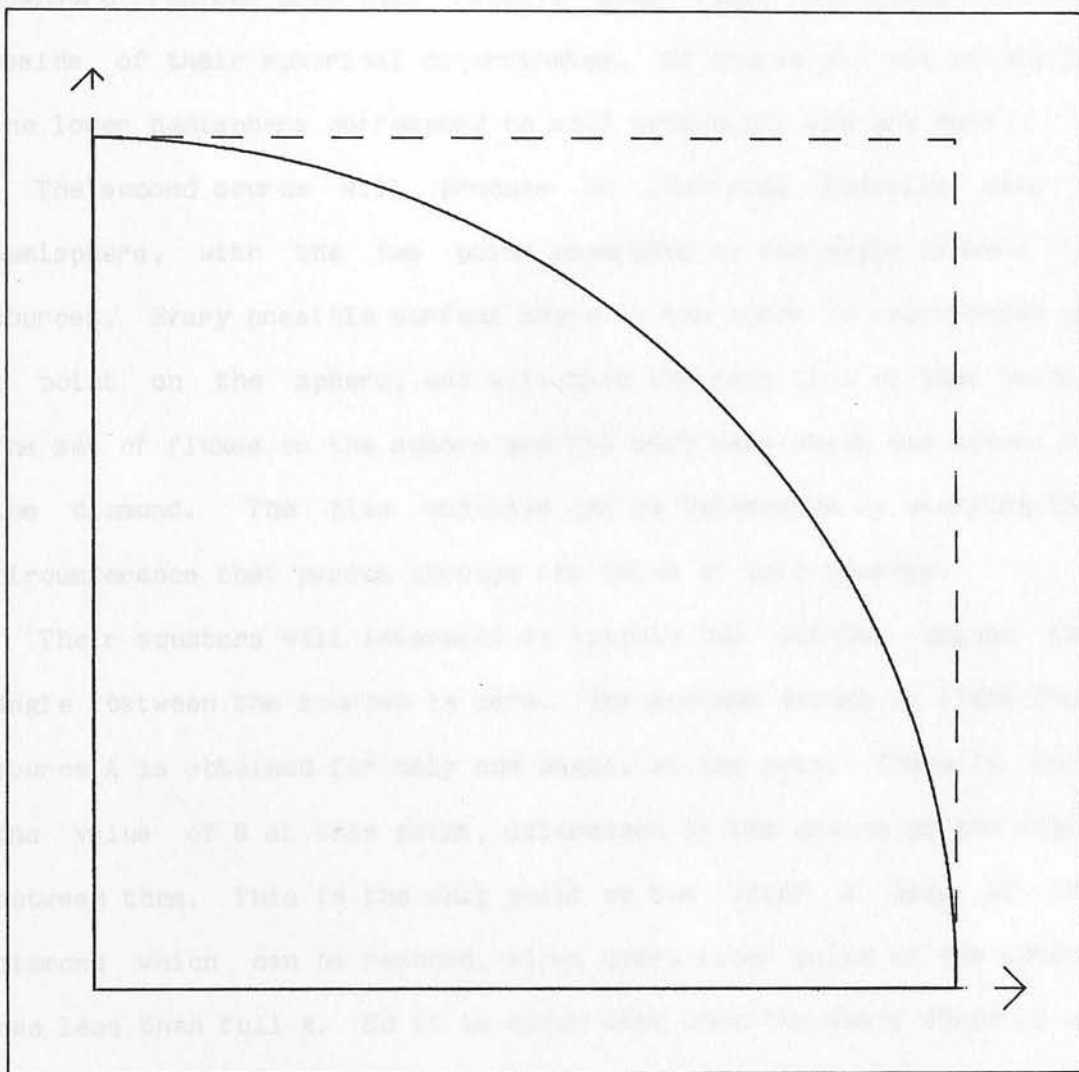
4.2 Envelope of angle .5 π radians

also the natural way to normalise envelopes to remove the effects of
the width of the diamond before subsequent processing. Diagram 4.2
shows the envelope characteristic of a source angle of .5 radians, or
90 degrees. Diagram 4.3 shows the envelope characteristic of a

source angle of .9π radians, or 162 degrees.

It is easiest to see the shape of the envelope by imagining a sphere in the scene. Say that the first source is directly overhead; the top of the sphere will receive the maximum flux. The equator of the sphere will receive no flux, and all the rings of latitude downward from the pole will receive less light according to the cosine of their spherical co-ordinates. Of course all the points in the lower hemisphere correspond to self occlusion, and are dark.

The second source will produce an identical function over a hemisphere, with the two poles separated by the angle between the sources. Every possible surface angle in the scene is represented by a point on the sphere, and will have the same flux as that point. The set of fluxes on the sphere are the only ones which can appear in the diamond. The flux envelope can be determined by studying the circumference that passes through the poles of both sources.

Their equators will intersect in exactly two points, unless the angle between the sources is zero. The maximum amount of light from source A is obtained for only one angle, at the pole. There is only one value of B at this point, determined by the cosine of the angle between them. This is the only point on the upper A edge of the diamond which can be reached, since every other point of the sphere has less than full A. So it is clear that when the exact shape of an envelope is determined empirically, that its maximum point in the direction of A specifies the angle between the two sources.

Similarly, the only flux value to lie on the upper B edge of the diamond corresponds to the pole of the B hemisphere. But notice that the amount of A here is the same as the amount of B at the A pole, since the angle between them is the same. In general it will be

found that the envelope is symmetric, after accounting for any
difference in brightness between the two sources. That is, if the
flux diamond is rescaled so that A and B are the same length, then
the envelope is symmetric around the main diagonal.

The brightest point in the main diagonal comes from the sphere at
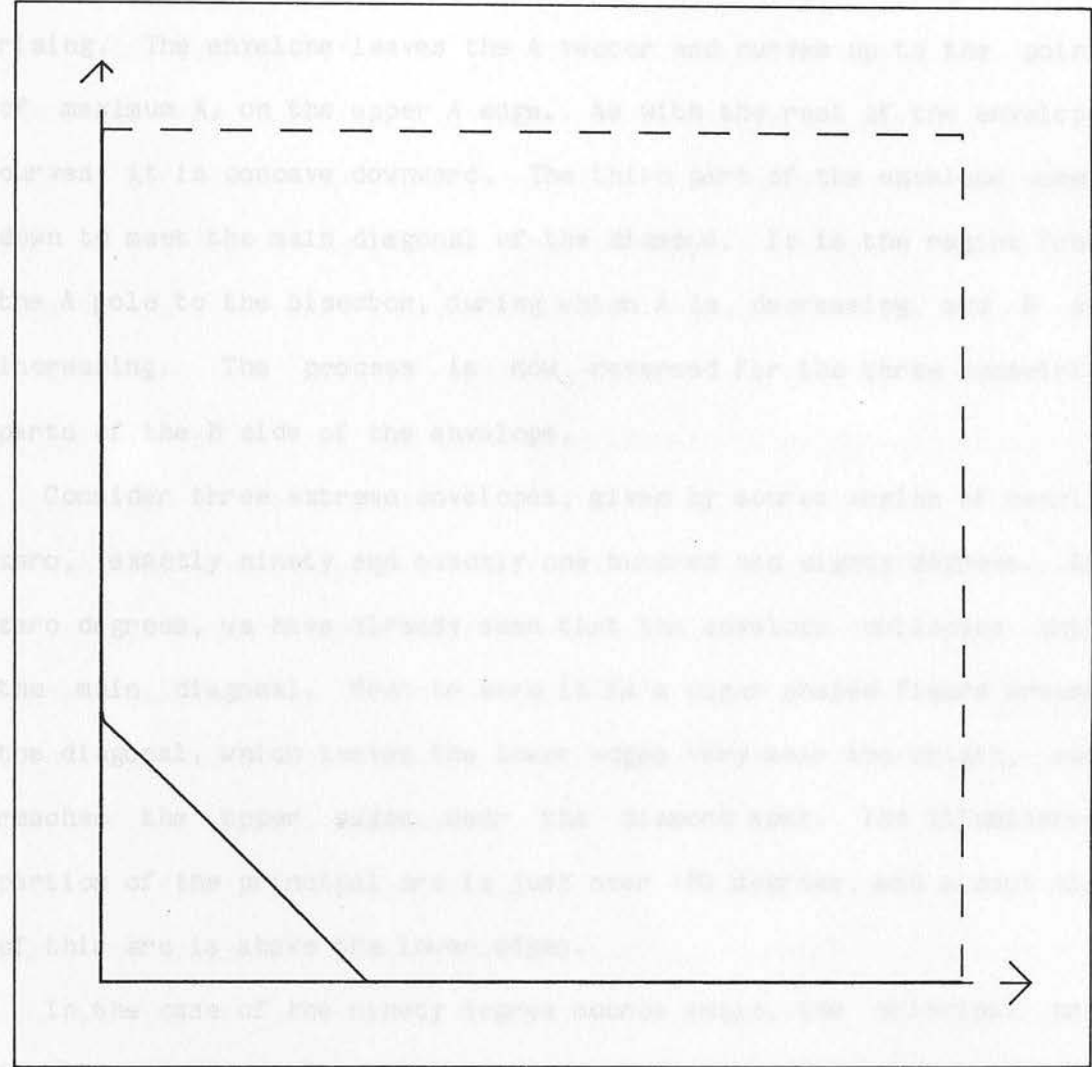
4.3 Envelope of angle .9 radians

the mid point between the source poles. The circumference which
bisects the two poles is the set of all the points where A and B are
equal, and so these are the points of the main diagonal. The
brightest one is the one at the top, between the two poles.

- 124 -

The envelope can be traced along the perpendicular circumference, which passes through both poles. We will call this the principal arc on the sphere. From the equator of A it rises through values of A until the equator of B is reached at a point determined by the angle between the sources. This first part of the envelope is on the A vector. The second part continues to the A pole, with both A and B rising. The envelope leaves the A vector and curves up to the point of maximum A, on the upper A edge. As with the rest of the envelope curves, it is concave downward. The third part of the envelope comes down to meet the main diagonal of the diamond. It is the region from the A pole to the bisector, during which A is decreasing and B is increasing. The process is now reversed for the three symmetric parts of the B side of the envelope.

Consider three extreme envelopes, given by source angles of nearly zero, exactly ninety and exactly one hundred and eighty degrees. At zero degrees, we have already seen that the envelope collapses onto the main diagonal. Near to zero it is a cigar shaped figure around the diagonal, which leaves the lower edges very near the origin, and reaches the upper edges near the diamond apex. The illuminated portion of the principal arc is just over 180 degrees, and almost all of this arc is above the lower edges.

In the case of the ninety degree source angle, the principal arc is 270 degrees. The equator of one hemisphere falls on the pole of the other, and so the envelope leaves the lower edge exactly at its end; the maximum value of A lies on the A corner of the diamond. The diamond swings in a circular arc to the opposite corner. Ninety degrees of the principal arc appears on each of the two lower edges, with the remaining ninety in the curve between the corners.

Between zero and ninety degrees the point where the envelope leaves the lower diamond edge gradually rises, while the point where it meets the upper edge descends, until they become the same at the corner when ninety degrees is reached. The cigar is progressively fattened. At forty five degrees the envelope equally rounds off the three non-origin corners.

At 180 degrees the two hemispheres intersect only at their equators, and this is the only part of the sphere which is not in light. Yet nowhere do both sources arrive at once. The principal arc is 360 degrees long. The first ninety degrees go up the lower A edge to the A corner. The second ninety degrees come back down the A vector to the origin. The remaining arc goes up and back the B vector. The only points that are within the envelope are the two lower edges of the diamond.

When the source angle is near to 180, the envelope does not come quite back to the origin after traveling up the A vector. Instead it makes a shallow arc, concave downward, to the B vector, just bypassing the origin. As the source angle progresses from ninety to 180 degrees, the arc between the A and B vectors moves down from the tips of the vectors to the origin, and becomes shallower. In each case, the brightest point of each source has no light from the other, and so the A and B peaks of the envelope occur at the two outside corners of the diamond.

Thus we have seen the entire range of possible envelopes, but so far we have been considering only the brightness effects due to surface angle relative to the sources. The other brightness effect which can occur in this domain of scenes is a shadow, which occludes one or more sources entirely. A shadow gives zero flux in the

- 126 -

occluded source, so will cause the flux point to lie on one of the lower edges, and any point on the full length of the lower edge is obtainable by some shadow. Therefore, when accounting for shadows, the complete flux envelope must also include the entire vectors as well. These will be called the shadow spikes. For source angles of less than ninety degrees, the upper part of the vector is not obtainable except in shadow. This means that if the source angle is known, then fluxes in the upper part of the lower edge of the diamond are known to be in shadow, fluxes off the lower edge of the diamond are known not to be in shadow, and only the small set of fluxes in the lower part of the lower edge are ambiguous.

When a shadow occurs with a source angle of ninety degrees or more, it still falls within the envelope, which already includes the entire lower edges of the diamond. This leads to further ambiguity, because even without shadows over half the surface of the sphere corresponds to points on the diamond edges. On the other hand, these envelopes due to obtuse angles are easy to recognise, since the edge spikes are noticeably higher than any point in the middle of the diamond.

The shape of the envelope is controlled by a second parameter in addition to source angle. The ratio of the brightnesses of the two sources, A : B, causes the envelope to stretch along one of the axes. For example, the envelope of diagram 4.1 has the same angle as the envelope in diagram 4.4, but the latter has a brightness ratio of 2:1. The same difference is true of diagrams 4.2 and 4.5. When the co-ordinates are scaled by this ratio, then the symmetric envelope characteristic to the source angle is obtained.

Since the two envelope parameters are constant over the scene,

their selection can be done by a best fit algorithm. If a region is well curved, then it occupies a substantial portion of the area under the envelope. By fitting the closest possible envelope around the flux points it is possible to estimate both the source angle and the source brightness ratio. This is best done over the regions
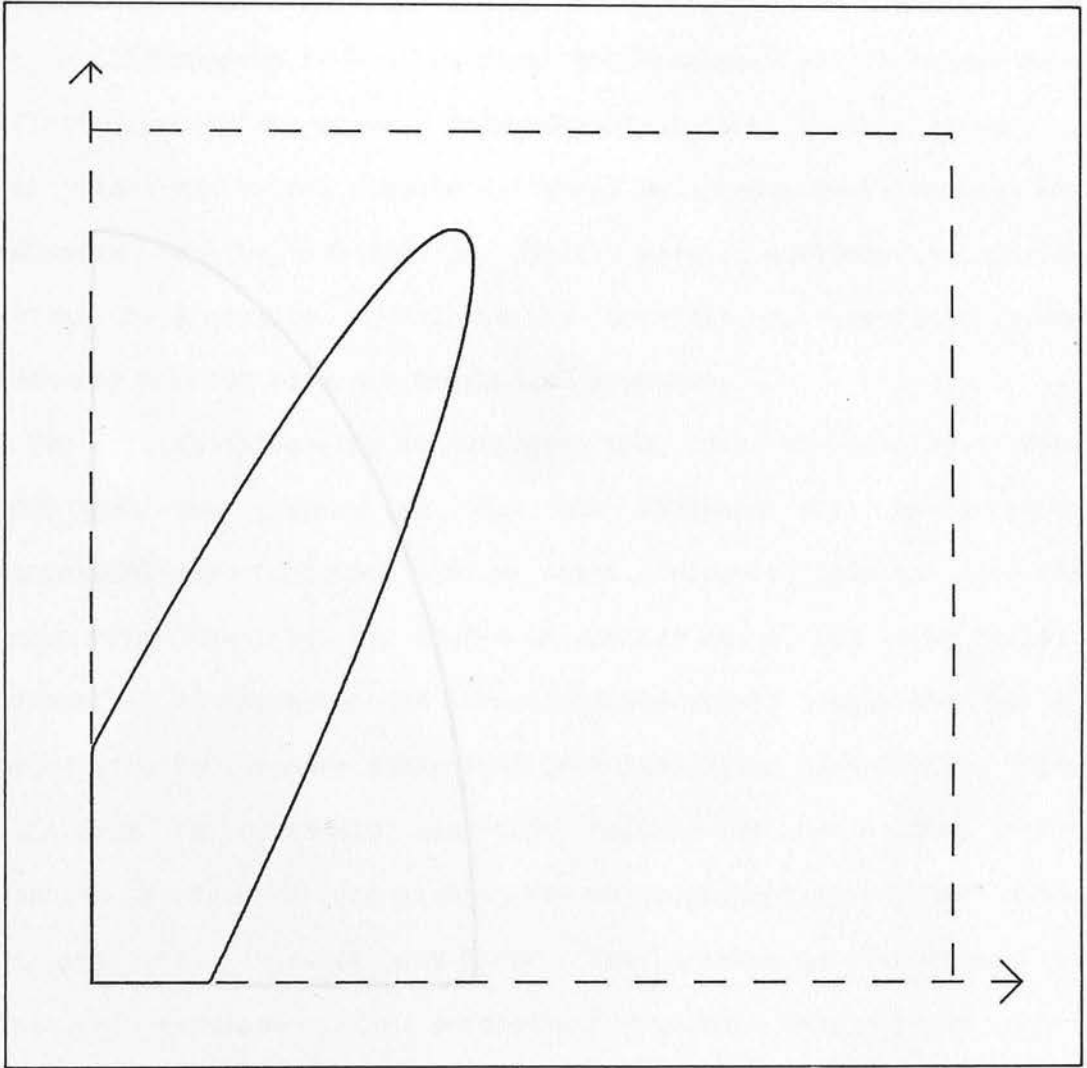
4.4 Envelope of angle .1 and ratio 2:1

collectively, since the single envelope must fit all the regions.

Conversely, the envelope may already be known, perhaps from a previous picture, or because the environment is controlled, or because it has been measured elsewhere in the scene. If the envelope

is known, then it can refine the brightness measurement of most of

the regions. Using pre-envelope diamonds, the brightness of a flux

region was simply taken as the greatest flux in A and the greatest

flux in B. However, we have seen that in this world, the maximum

value of A can only occur at a single angle of surface, and so is a
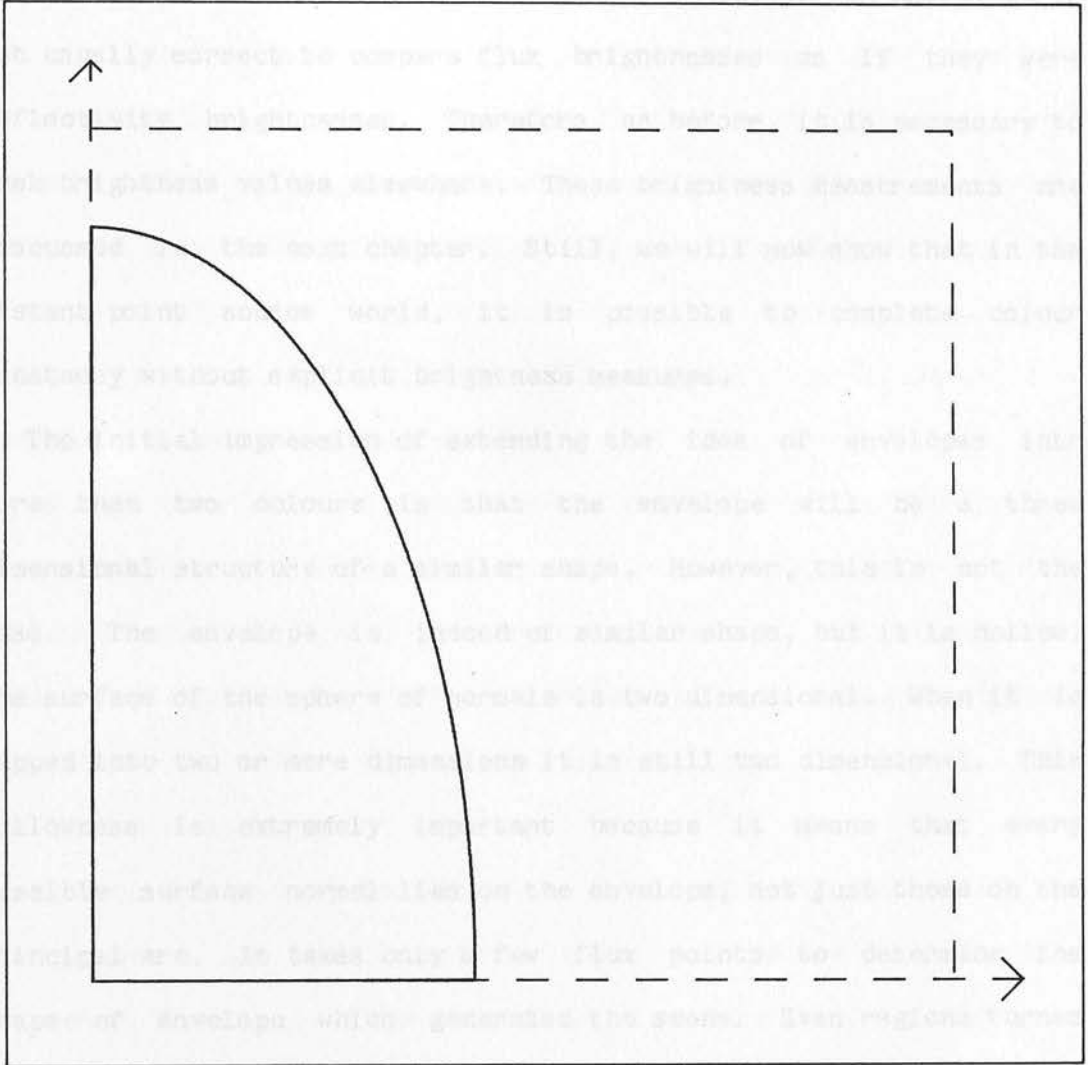
4.5 Envelope of angle .5 and ratio 2:1

rare occurrence. A better method is to fit the smallest envelope, of

the known shape, to the flux points, and take its maximum flux in

each direction. In a sense, the previous, non-envelope, method was

simply fitting the smallest diamond-shaped envelope to the data.

In two dimensions, this technique suffers the same basic flaw as the non-envelope brightness estimates. If the region happens to give entirely low flux, it is not clear whether this is due to it being a dark reflectivity, or due to not getting as much illumination. If the surface normal never crosses the principal arc, then the envelope fitted will not represent its correct brightness. In short, it is not usually correct to compare flux brightnesses as if they were reflectivity brightnesses. Therefore, as before, it is necessary to seek brightness values elsewhere. These brightness measurements are discussed in the next chapter. Still, we will now show that in the distant point source world, it is possible to complete colour constancy without explicit brightness measures.

The initial impression of extending the idea of envelopes into more than two colours is that the envelope will be a three dimensional structure of a similar shape. However, this is not the case. The envelope is indeed of similar shape, but it is hollow. The surface of the sphere of normals is two dimensional. When it is mapped into two or more dimensions it is still two dimensional. This hollowness is extremely important because it means that every possible surface normal lies on the envelope, not just those on the principal arc. It takes only a few flux points to determine the shape of envelope which generated the scene. Even regions turned nearly perpendicular to the sources and so having rather dark flux still specify an envelope, which in turn indicates what flux this region would have if it directly faced one of the sources. In short, it is not possible to fit an envelope which is too small or too large to the flux data, so brightness information is contained in every envelope.

A comparison of the envelopes is the same as a comparison of the regions' reflectivities. Smaller envelopes imply darker reflectances. The envelopes are located at different angles from the origin according to the same transformation as their parallelepiped, and so the reflectance colour ratios can be measured.

To complete colour constancy it is only necessary to eliminate the bias of the illumination colour and brightness. Any colour cube technique, discussed in the following chapter, will suffice. Land's method would find the component-wise maximum over all the envelopes, define it as virtual white and factor it out. The body-fit algorithm would consider the range of envelopes, fit a circle around their collective middle, perpendicular to the diagonal, and define white in terms of the centre of this circle.

In this case of three or more source colours, if shadows reduce the illumination at some pixel to only two sources, then the flux is projected onto one face of the parallelepiped. If the envelope is known for that region, then this leaves the pixel ambiguous between two values. It would often be the case that local information would resolve this ambiguity.

It should be noted that some of the envelope uniqueness in a three source scene is lost when the three sources are co-planar with respect to the scene. From the sphere of normals it is clear that the principal arc through the sources divides the sphere into symmetric hemispheres. Therefore, off the principal arc any given flux point can represent either of two surface normals. Disambiguating these would generally require additional knowledge, as in the two colour case. However, colour constancy would still be possible, because every pixel still falls on, rather than inside, the

envelope, which curves in three dimensions as before. The change in the envelope is that it has collapsed into a curved plane, rather than a "balloon". The completion of colour constancy requires only that brightness is measurable in many regions. This in turn·requires that these regions have points on the envelope, which is easy since every point is still on the envelope, and that the location relative to the envelope is measurable, which is satisfied because the envelope curves.

Returning to the general case of three or more independent sources, it seems that the lighting configuration requirements can be relaxed without loss of the ability to perform complete colour constancy. Recall that the reason for making the point sources distant is to eliminate the changes of brightness due to distance from source. However, notice an important distinction between distance brightness and surface normal brightness. The surface normal may change arbitrarily between pixels within a region, but the distance from a source can change no more than the distance between the pixels. So the distance brightness function is gradual within each region, and roughly systematic over an image. Therefore, it is possible to relax the distant source requirement, and extract the distance function after finding the surface normals. The first stage is estimate the envelope locally for various regions; then use the error between the actual flux and the ideal envelope to determine the distance function; finally, compensate for the distance changes and recompute the envelope more precisely.

The distance function can be measured in two ways: across the scene and within each region. The first assumes that the function is reasonably smooth across the scene. It takes the proportions of the

parallelepiped defined by the envelope at many diverse regions. These proportions will change gradually over the scene, according to distance from each source. Normalising them will coarsely remove distance brightness. The more regions which were measured, then the less interpolation between regions, and so the greater the accuracy. This method has the drawback that each region is treated as having a constant distance brightness, which becomes less true the larger the region.

The second method is to isolate the small brightness changes within a given region. Having estimated the envelope, subtract its value from the flux, which should leave a smooth error function, attributable to local distance changes. Generally this would work better on larger regions. The overall brightness compensation would be most accurate if both methods were used together.

Notice that the normals and distance functions are not completely independent, since as the normal changes, the region's rate of approach to the source will also change. Compensating for this will produce a more accurate algorithm. It should be well behaved because the dependence is constant, up to any parallax effects, across the entire scene.

It should be noted that this general colour constancy algorithm for scenes under three or more independent sources, with or without the distance brightness compensation, uses no edge measurements at all. The surface normals have relatively little inherent connection with location in the scene. This puts them in the domain of flux space techniques, where the attributes of a set of pixels are considered without reference to scene position. The envelope does not contain any more scene-spatial information than the fact that the

flux values all come from the same region.

Brightness due to distance functions are not in this class, because they are inherently tied to the distance from the sources within each region. Indeed, this distinction is the fundamental reason that the two effects can be separated.

Throughout this development the surface normals have been treated as a by-product. The interesting characteristic has been that the set of possible normals restricts the set of possible points in the flux structure. However, the normals themselves are quite useful. It is well known that scene depth information is obtainable from the surface normals. This is because the normal is a good predictor of the change in depth to the next pixel. By interpolating the normals, a good estimate of the third dimension of each region may be obtained.

Further, some object recognition algorithms use the normals rather than the dimensions of the object. Notice that the changes in the normals are object-relative attributes, and so more useful for object identification and measurement than scene or imager-relative attributes.

Envelope theory has implications for how to select source angles and brightnesses when the illumination can be controlled, as in an industrial setting. In discussing these, we will assume that there is only one camera, of fixed location directly above the scene, so giving a view of one hemisphere of the sphere of normals. If the camera is not perpendicular to the support surface, then some of the theoretically visible normals will be obscured by contact with the floor, and so less than a hemisphere will be visible. Of course it

would be impossible to see more than the hemisphere without multiple cameras.

Likewise, it is assumed that all of the sources are of constant predetermined brightness, colour and position. We will not treat the cases where they may be switched on and off, or their brightness controlled during or between pictures. However, we will assume that brightness, colour and position can be set to any desired value.

Controlling the lighting gives several obvious advantages. Since both the source angles and their relative brightnesses are known, then the envelope shape is known. The angle between the camera and sources is known, so the surface normals can be converted into absolute co-ordinates from source co-ordinates. Beyond this, the brightness and location of the sources are known, and so the brightness-from-distance function is predictable when necessary.

In many cases, the choice of equal brightness of the sources, and so a symmetric envelope, will simplify computations by eliminating the divisions involved in re-scaling. However, the sensitivities of practical receptors, and sometimes the likely range of objects viewed, may force different choices. The receptors must be within their working ranges, neither saturated, nor at a low level where noise becomes more significant.

The choice of source angle involves a trade off. In general, having sources coming from an angle near the camera means that they illuminate the greatest portion of the visible hemisphere. A wide angle from the camera increases both self-occlusion of objects and the extent of visible shadows. However, if all the sources have a narrow angle to the camera, then they also have a narrow angle to each other. This means that a given change in normal causes a

proportionately smaller change in flux, and this reduces the accuracy of the measurements. A rough guide is to look at the area of the envelope that does not lie on the lower edges of the diamond. This forms the range of possible results. If the area is becoming very small because the sources are nearly co-linear then the flux differences between normals will be proportionately smaller. This phenomenon is aggravated by the fact that for narrow angles more of the sphere lies off the lower edges, so more points are packed into a smaller area. On the other hand, having more points off the diamond edge is generally good, since more normals are potentially illuminated by all of the sources.

Notice that while shadows are usually undesirable in envelope calculations, if photometric stereo algorithms are being used as well, then the shadows may be regarded as a positive attribute. In this case, the considerations of the stereo algorithm would generally control the angles of the sources.

One possible solution to the problem of the sources clustering around the light is to have two opposing sources of the same colour. Additional sources of the same colour are much less expensive than additional colours of sources because the latter generally necessitates an additional receptor channel. The two sources of the same colour would be 180 degrees in opposition, so that all of the sphere is illuminated, except their mutual equator, and yet no point receives light from both sources. Normally the sources are both perpendicular to the camera. We will call a pair of this type a double opposing source. An advantage is that it gives strong illumination to areas turned well away from the camera, which are often dim and thus hard to measure accurately.

A second unusual type of source is co-linear with the camera. This could be implemented through the camera lens by an image divider, or else from a ring source just around the lens. It illuminates exactly the hemisphere seen by the camera and is not susceptible to shadows. Further, there is no computation involved in converting the surface normal from source co-ordinate to camera co-ordinate. We will call this a camera source.

One reasonable configuration of sources is three sources near the camera, with perhaps forty five degrees between them. If they are in an equilateral triangle, then there is an elegant symmetry, both of illumination coverage and of computation. If they form a T, with two on opposite sides of the camera and the third at right angles to this line, then symmetry is reduced, but it is much easier to relate the lighting and surface normals to the rectangular co-ordinates of the scene. If the coverage of a camera source is desired, then a similar result to the T would come by adding the second and third sources at right angles to each other, forming a V with the camera source. As with other uses of the camera source, it pushes the remaining sources further from the camera for the same amount of flux resolution.

Remaining with three sources, but using a double opposing source, there are several reasonable possibilities. Notice that if a double opposing is used with a camera source, or if two double opposing sources are used at right angles, then there are now two symmetries on the sphere and thus four possible normals for most flux values. The third source must distinguish between them, and this is only done if its plane is not a multiple of forty five degrees from the double opposing source. If there is no camera source and only one double opposing, then one or both of the remaining sources should be placed

out of ninety degree alignment with the double opposing axis.

When four sources are available, there is more freedom of placement. Using the four in a square around the camera, with perhaps forty five degrees scene angle between the opposite ones would have several advantages. It is symmetric and converts easily to rectangular scene co-ordinates. Every point on the visible hemisphere receives light from at least three sources, and shadows are only likely to affect one source, again leaving three.

Three sources might seem natural with a conventional colour television camera, which has three receptor channels. However, it would be easy to get more receptor channels. If the scene is static, a mechanism could rotate any number of filters in front of a monochrome camera. If the scene is to be taken in one picture, then a beam splitter can divide the scene through filters to separate monochrome cameras. This involves weakening of signal, alignment difficulties and expensive equipment. An alternative is to use a CCD camera, and prepare a mask which places a different colour filter over each pixel in a systematic pattern. The natural numbers of receptors are squares of integers: 1, 4, 9, but any number would be possible. The masks would be inexpensive to mass-produce, but this method means that there are less pixels per colour, so spatial resolution is reduced from the already small size of the contemporary CCD camera. There is also a difficulty with the alignment of the colour planes, since each pixel is sampling a unique location; interpolation or similar processing would be necessary.

We will now consider the results of implementing flux envelope theory in a computer program. There were in fact four separate programs. Two worked with two-source envelopes, and two with three-

source envelopes. In each case, one program measured the size of envelope appropriate to a region, given that the correct envelope is already known. The other program was to find the best fit of envelope to the flux points of a region, and so determine the correct envelope for the scene. A pre-processor transformed all flux structures into a standard Cartesian plane or space, removing unnecessary complexity. None the less, the problem is much more complicated than it seems to a human who can immediately provide a good estimate of the envelope from a visual inspection of the flux points.

Two dimensional envelope measuring consists of finding the smallest envelope of a given shape which will enclose the data points. It is functionally equivalent to finding the upper edges of a flux diamond. The method used was to define a standard sized envelope of the given type. A function was written which returns the distance of this standard envelope from the origin for any angle. This function could compute the trigonometric solution, or use table lookup and interpolation, or use a fast one-dimensional climbing algorithm, which was the method chosen. Now for each flux point in the region, compute the angle and distance from source. For this angle use the first function to determine the standard distance. The ratio of actual to standard distance is the proportionate size of an envelope which passes through this point. The ratio can of course be larger or smaller than 1. The largest such ratio among the region's points is the region ratio. The desired envelope is the standard envelope scaled by this ratio, but in general it is the ratios themselves which are used in further processing. Notice that it may not be necessary to test every pixel. Only those on the

convex hull are candidates. The algorithm also correctly processes points on the edges of the flux structure. When the source angle is less than ninety degrees there is provision for separate processing of shadow and non-shadow edge points, because there is a band of edge points which should only be obtainable in shadow.

The three dimensional envelope measuring was substantially different. Several parts were much more complicated. In particular the standard envelope function was climbing in a two-dimensional space, where it is able to get lost or trapped by local maxima. It was no longer a case of finding the largest fit, because all the points in a region should be on the same size of envelope surface. Therefore each ratio was tested against the others within a tolerance band, with manually processed error reporting. Although a single point is sufficient to define the envelope scale for the region, once the envelope shape is known, the thorough testing was considered more appropriate for an experimental context.

Envelope fitting algorithms, which find the best shape of envelope to fit the data, involve more theoretical issues. There are two cases for the two-source fit - curved objects and polygonal. If the scene has regions of substantial curvature, then a region is almost certain to curve across the principal arc. This is noticeable as a local maximum in one source colour associated with a characteristic spatial gradient. The virtue of detecting these cases is that then it is known that these points lie on, rather than merely not above, the envelope. Processing of this particular feature has not yet been implemented.

The polygonal case of two-source fit has been built. The data points could come from curved objects, but the method does not

utilise this knowledge. The approximate algorithm was to fit a range of envelopes to the points, using the envelope measure algorithm above, and then decide which is the best fit. In fact any envelope shape will enclose the data if it is large enough. The problem comes in finding an appropriate way to compare best-fit envelopes of entirely different shapes. It turned out that beyond ninety degrees of source angle, when the axes are longer than the middle of the envelope and the middle approaches a circle, that a good measure is the length and ratio of the axes themselves, tempered with the maximum distance to any non-edge point. In any event, these cases are less important that acute angled sources.

To compare envelopes, several measures were tried. The distance to the centre of the envelope tends to give envelopes that are too fat, thereby overestimating the angle between sources. The distance to the corner of the enclosing diamond reduced but did not eliminate this effect. Since the envelope is defined by two parameters, source brightness ratio as well as angle, the measure should also correctly compare envelopes of different ratio. One measure tried was the area under the envelope curve, but it fails dramatically. The area of the enclosing diamond gave some of the best results of those measures tried. Typically the angle between sources would be found to better than fifteen degrees and the ratio to better than twenty-five percent, given only six random surface angles in the region. These results are poor in comparison with the accuracy of colour constancy, but none the less they give a new way of roughly determining source angle.

In retrospect, the important characteristic of the points for fitting envelopes is probably not the presence of points near the

furthest point of the envelope, but the absence of points low down, where the envelope meets the axis. These might be called the notches. The next fitting algorithm will probably define a measure for the proportional depth of the notches and use this for the angle. The ratio would be a combination of the ratio between the two notches, the ratio of the enclosing diamond, and the angle to the most distant points.

Surprisingly, fitting an envelope to three-source data is easier, because there is only one envelope that can fit. There is no need to find a way to compare envelopes. The envelope is defined by five parameters, one for the angle between the first two sources, two more for the angle of the third source relative to the coordinates established by the first two, one for the ratio of the second source to the first, and one for the ratio of the third source to the first. It was assumed that at least five different data points would be available in some region. The five parameters define a five-dimensional solution space. The method used took the first point and found some envelope that fit it. It then successively added points, tracing a surface in the solution space so that the previous points still fit, as did the new point. This method was sufficient to show that the envelope can be found from a group of points, but it is not satisfactory on theoretical grounds. The nature of the solution space has not been adequately explored to show that climbing will reach the solution for all envelopes and all data points. Further, it is not even known what independence conditions are sufficient for five data points to be sufficient to define a unique envelope. If three-source envelopes are to be utilised further, their mathematics must be studied in greater detail.

This chapter has looked at the limited domain of scenes illuminated only by point sources, with only one source for each colour of illumination. By studying the effects of the illumination on the scene, it has been shown that there is a strong relationship between the angle of the surface to the lights and the amount of flux which is seen from it. This allows the surface normals to be predicted in such scenes with a high degree of confidence. For scenes of mainly curved rather than angular objects, the surface normals can be used to provide depth information and object identification. Thus for a particular class of scenes we have derived most of the information about the physical geometry of the scene, using only the changes in colour within the regions, and at no time using edge detectors or measurements.

This domain of scenes has an additional property that in no case can all the points of the flux diamond be occupied. The set of points which can occur is determined by the angle between the sources, and is called the envelope. If the angle between sources is known, then the envelope is determined, and knowledge of it can be used to refine the other colour constancy algorithms. On the other hand, if it is not known, then studying the flux patterns within the diamonds will allow it to be estimated. Similarly, the relative brightnesses of the two sources is directly related to the shape of the envelope, and so one can be determined from the other.

When three or more sources of independent colour occur, then every surface normal has a distinct position on the envelope, and brightness comparisons can easily be made between surfaces, thus completing colour constancy. The information is strong enough to be also able to compensate for brightness changes due to distance from

source. This produces the slightly surprising result of a colour constancy algorithm which can handle general three dimensional scenes, but only if the scene is complicated with three or more sources.

These computations with the flux envelope and the surface normals are applicable to scenes with both controlled and uncontrolled lighting. Under uncontrolled lighting it is possible to determine several properties of the lights, including their relative brightness and their angle to each other and to many points in the scene, in addition to the information available about the objects in the scene. However, the algorithms are more likely to be used under controlled lighting, which is selected to yield the most information about the scene given the class of objects that will be viewed. We have seen some of the considerations in choosing effective lighting.

Chapter Five

Scaling the Colour Cube under Single and Multiple Illuminants

Edge Types and Brightness Measurement

Utilising Relative Brightness Information

Critical Brightness Ratios

Land's great contribution to creating colour constancy algorithms
is the notion of the colour cube. This chapter develops the cube to
take advantage of the information extracted from a scene by the flux
space methods. In particular, the ability to separate the effects of
multiple illumination colours gives rise to a colour cube for each
source colour.   The interaction between these cubes can refine the
overall perception of white, and thus come closer to interpreting the
correct surface colours than would result from the same scene with
less source colours.

We saw in chapter two that Land's method of edge detection and
measurement was of little use outside the Mondrian world, and indeed
it was even more primitive than many monochrome edge detectors. This
is bearing in mind that he made no claims for its ability in three
dimensional scenes. In chapter three we provided an alternative
region finder which was truly colour based. In all but the most
degenerate circumstances it gives the colour ratio between any two
regions.   However, so far it does not give the brightness ratios.
Techniques for measuring and using brightness ratios will   be
considered in this chapter. They involve links between the spatial
arrangement of pixels in the scene and the otherwise spatially
independent flux space structures. A method is given for identifying
exactly those edges which have the same illumination on both sides,

- 145 -

and so have a reliable brightness ratio.

Having accumulated a group of regions for which the brightness ratios are known, a colour cube is fitted to them in order to complete the colour constancy. Land's method of scaling the cube was by computing the component-wise maximum of the regions, which we have been calling the virtual white. An alternative scaling, called the body-fit, is proposed. It uses information from more regions, and so may be more reliable, although a combination of the two methods is probably best.

Finally there is the notion of critical regions and critical bands of brightness ratio. These allow selective attention when determining brightness ratios in a scene.

We will first consider a procedure for doing brightness compensation without using edge techniques. Like other flux space techniques, it works better as great numbers of sources illuminate the scene. In a reasonable scene, without contrived gradients or strong alignment of surface normals, under two or more sources it should give good estimates of the brightnesses, and so complete colour constancy. As scenes have more coincidental alignments of gradients, edge measurements such as those given later can be added to improve the result. The first two stages rely on the fact that brightness gradients are usually gradual and systematic over a scene. So if a source is to the right of the scene, regions of the scene to the right tend to get more light. For three dimensional scenes, the function is not perfect from the observer's viewpoint, but there is still a strong trend. It is assumed that the object colours and orientations are not contrived to simulate this behaviour. Notice that scenes with distant illumination, such as in the previous

chapter, will not need the first two steps, nor will those steps alter their value.

The first stage of the procedure occurs only if there are multiple sources. It attempts to compensate for any difference between their brightness gradients. For example, a red light comes from the right of the scene and a blue from the left. If a sample of flux structures is taken from various parts of the scene, it is observed that the diamonds change shape. To the left the blue axis is proportionately much longer, and the reverse is true on the right, while the centre is between the extremes. This observation is not biased by large red objects for two reasons. Firstly, each object produces a single diamond, irrespective of its size, so large objects do not have more effect than small objects. Secondly, it is not a comparison of the amount of flux in two colours, but the ratio of two sources on the same region; if the diamond from a red surface shows that the red source is twice as strong, then a blue surface will show the same thing, only with the diamond transformed sideways a bit in flux space. The only way to fool this algorithm is to align the normals so that an area has regions which mostly point towards one source in preference to another, and even this case can be detected with better envelope techniques. Knowing the ratio of source brightnesses in different parts of the scene allows changes in these to be equated, so that the ratio is constant across the scene. There may still be a collective gradient where one side of the scene gets more light than the other, but this is left for the next stage.

The second stage removes a single common gradient, whether there is only one source, or several which have been processed as above. One heuristic is to look for objects of the same colour around the

scene. This is easy because they will be coplanar in flux space. They might be leaves or flowers, or any object which is common in the scene. These would usually have the same brightness as well as colour, so a decrease in flux suggests a decrease in available light. A second heuristic is to collect the brightest and darkest fluxes from each region in an area, with surfaces of the same colour counted as a single region. If the range of these fluxes is low, again it suggests that the region is dark. A large area of a dark or light object will not bias this test since it counts as only one region. Regions which face mainly toward or away from the light would produce a narrow range between their darkest and lightest pixel, and so would be weighted against. When multiple source colours were present, it would be more difficult to fool this technique because a change of normal towards one light is a change away from another. It is of course possible to confuse it by placing all the dark objects on one side and all the light objects on the other. Having estimated the brightness gradient across the scene, the values can be altered to compensate for it. After this stage there should be no large effects of brightness gradient in the scene by any of the sources.

The third step applies only to scenes with multiple source colours. It allows the brightness found under any of the sources to be transferred to the others. In the first step, the brightness ratios of the sources were determined for separate areas of the scene, and then made constant. This ratio now has another use, in determining the proportions of the flux structure. If there are two source colours then the structure will be a parallelogram, or diamond. The ratio of two adjacent sides of the diamond will be the same as the ratio of the source brightnesses. Previously, when given

flux data in the flux surface, the smallest enclosing flux structure was chosen. In the case of a diamond, the upper edges pass through the max A and max B points respectively. This can produce a diamond of nearly any shape. Given the true brightness ratio of A : B we can do better than this, since we know the shape which every diamond actually has. The true flux structure is then some diamond of this shape that encloses the data points. In particular, it is at least as large as the smallest enclosing diamond of that shape. If, for this region, max A and max B happen to have the same ratio as the sources, this method will produce the same flux structure. In all other cases, one of the two upper edges will be extended, and only one of the region's maximums will be on an edge. The actual result of this algorithm is that if say region-max A is 90% of scene-max A, while region-max B is only 70%, we compute the flux diamond as if both maximums were at 90%, because if the surface faced source B, at least this much flux would result. We can not further enlarge the flux structure by this algorithm because it only measures how much one source colour lags behind the other in a region, rather than how much they jointly fall short of the scene maximum.

A more succinct way to look at this technique is that it applies a rectangular shaped envelope to the scene. The proportions of the envelope are taken from the brightness ratio of the sources. It is known that all the regions of the scene potentially could get this ratio of the sources, so if it does not occur it must be because some source has gotten less than the maximum, rather than the other getting more.

This can be very powerful. By definition some pixel in the scene receives max A. The region-wise corrections automatically compute

the value of max B in the same region. This also works in the opposite direction. Without this ability, there is an independent colour cube for each source colour. With it, each bit of information in one colour is immediately transferred to the other cube. Indeed, computationally it is then easiest to compute only one cube over the scene, by taking a single brightness value for each region. This is better than the single cube generated by a scene with only a single illuminant, because now two different sources have a chance to illuminate each region. The brighter of the two is always taken. If source B was not useful, then max A would be proportionately greater than max B in every region, which is very unlikely. Additional source colours would typically continue to improve the cube, but their marginal benefit would usually decrease arithmetically. Dispersing the positions of the sources to illuminate thoroughly each scene portion and surface angle, should maximise the effect.

Now if it is not possible to find out the relative source brightness before preparing the colour cube, it is still possible to make a refinement of the colour cube similar to this third step. The technique to be considered now is not a perfect algorithm, in that it is possible for it to generate a slightly incorrect answer, unlike the preceding technique. On the other hand, it does not require any information beyond the basic flux structures which are combined into the cube. If the preceding algorithm is used, then this one will have no further effect on the result.

The most straight forward way to process the flux structures is to separate the effects of each source, and compute a separate flux cube for each illumination colour. Recall that the upper corner of a cube, called virtual white, represents the least component-wise value

- 150 -

which a white surface would show under the maximum amount of light of that colour in the scene. In fact the actual white can be any flux which is at least this big in each component. Geometrically, in the three dimensional case the locus of possible actual white points forms the positive octant of the space, with the vertex at virtual white. Since as a point moves away from virtual white within this octant, it is also substantially moving away from virtual white, deviations from the virtual white have comparatively small cost; a large movement from virtual white gives a much smaller movement in the vector's angle with the origin.

Now with multiple source colours, each colour has its own cube, and each cube has its own virtual white and accompanying octant. The set of actual whites, one for each source colour, will lie one in each respective octant. However, we know something else about the actual whites. They will bear the same relationship as the sources in the flux structures. In two colours this is the angle between the two edges of the diamond. If one region in the scene happens to have the highest flux in both sources, then it will determine the virtual white for both cubes at this stage, and so the virtual whites will be aligned. However, it is more likely that the cubes are composite, with different regions being maximum in different components, and so the virtual whites are unlikely to hold the true and known colour ratio. The virtual whites can be improved by increasing them to points in the octants which bear this ratio. This is the inexact portion of this technique. Normally the points chosen will be the ones which minimise the collective distance from the first virtual white points. Then new points will all be on the surface of the old octant, and for c components will cause each virtual white to

increase in between 1 and c-1 components, with the remaining components remaining constant. Each component will change in at least one source. The further the brightest regions are from a white reflectivity, the more effect this technique will have, but these are exactly the cases where it is needed most.

In some cases where the ratio of A : B strength is not known accurately enough initially, it might be reasonable to do the region-wise corrections of steps one and three on the basis of the brightness ratio computed after this virtual white adjustment has been made.

Virtual white is not the only way to scale the colour cube. It relies on the assumption that for each component and after any edge based transfer of brightness and colour, some pixel which is quite reflective in that component is also brightly lit in that component. Although this is frequently true, it only make use of the brightnesses of these few bright regions. We now propose a body-fit technique for scaling the colour cube. Its strongest advantage is that the cube can be colour scaled before the brightness compensations are done. This means that the colour of the illumination can be determined before the edges are measured for brightness, or any spatial understanding is attempted.

Examining Land's colour cube [1977, page 120] which has had the locations of actual colours marked in, it can be seen that the entire cube is not dense. Most of the points in the colour cube do not correspond to any naturally occurring colour. Instead, the set of all real colours is limited to an area close to the axis of the colour cube. The set tends to be thin near both black and white and a bit wider near the middle of the cube, forming a double cone. When

a cube has been accumulated from flux data, its points are almost certain to lie within this double cone.

This suggests the body-fit technique. Fit a circle around the middle of the actual colour cube and centre it in the ideal colour cube. The amount that this circle is smaller than a circle around the middle of the ideal cube indicates how much error is possible in this fit. So now it is possible to determine a confidence level during scaling. This gives the direction of the diagonal of the cube. The size of the cube could be determined by taking the centre of mass of the flux points. This would not be confused by large objects in the scene, but it would be biased by a large number of objects of similar colour. Instead, take either the furthest point in the direction of the diagonal to be the cube brightness, or do a general fit of the double cone and extrapolate the location of the upper cone point. It would be sensible to scale the cube by both the body-fit and the virtual white techniques and compare the results, which should be similar for well behaved scenes.

The body-fit can also be applied to flux space before it has become a colour cube by adding brightness information. Within the flux space for a given source, each region has a vector corresponding to the colour of flux caused by that source. After brightness scaling, the point in the colour cube corresponding to a region will lie on its vector. In other words, flux space is a colour cube which lacks brightness information. Now if a circular hull is fit around these flux vectors, it will closely approximate the colour of the source. From the colour of the source and the previously determined relative surface colours, the absolute surface colours are determined. All that remains is the brightness of the sources and

surfaces.

Up to this point in the thesis we have largely ignored edges.
Partially this is due to the excessive historical interest in edge
detectors, which reached the point of providing "interesting"
detectors which had no noticeable usefulness; frequently the solution
was announced without ever being able to find an appropriate problem.
Therefore, we have been interested to see how much useful colour
processing can be done without recourse to studying edges. However,
it has turned out that this was not just an attempt at a virtuoso
display. A fuller use of colour information than previously occurred
has lead to the discovery that the flux space approach is very useful
in colour processing, yielding up the number of colours of
illumination, separating the effects of each, and giving the colour
ratios between any scene regions. Although the flux space approach
is closer to region-based than edge-based, it is not truly region-
based because it ignores all spatial structure within the region
after the first step, when the separate flux planes are found.

Despite the usefulness of these methods, the present study has not
found a wholly satisfactory way to measure the brightness ratio
between regions in all types of scene. It can be approximated in
well behaved scenes with distant light, from the naive colour cube,
and better approximated by the various enhancements of this chapter.
It can be determined by envelope techniques for three or more
sources, and in favourable circumstances can be computed directly by
the secondary reflectance methods of the next chapter. But in
effect, we have reduced colour constancy to brightness constancy, and
have provided some help in solving brightness constancy. We will now
turn to colour-based edge measurement, to see how the information so

far obtained can be used to enhance monochrome brightness ratio measurements.

We saw in chapter two that in the Mondrian world the region brightness ratio was the same as the edge flux ratio. This is because all the regions are coplanar and there are no sharp shadows. Therefore, the difference in available light between two adjacent pixels is minimal. Outside the Mondrian world region boundaries are not so well behaved. If the region boundary forms a spatial edge, that is if the limit of the normal approaching the edge is different in the two regions, then the angle between surface and source will be different on the two sides of the edge, except due to accidental alignment. Thus the amount of light received from the sources will differ. In the case of an occlusion of objects from the point of view of the observer, the distance of the surface from the source will also change, and either or both regions may be in shadow relative to any of the sources.

The flux space techniques give the necessary information to distinguish between some of these cases. For example, having separated the flux effects due to each of the sources, shadows in any region can be recognised because there is zero flux in that particular source colour. A shadow in one region but not a neighbour is strong evidence of an occlusion between the objects. Note that it does not immediately say which object is closer to the observer, although there is some tendency for shadowed objects to be further away. A shadow edge which is common to both regions is good evidence that any depth occlusion between them is small.

Looking now at the surface edges under two or more sources, the ability to match the flux structure points after the transform is the

indication of the edge geometry and hence reliability. Constant flux on both sides of an edge would normally come from a distant source, lighting regions which are planar at their mutual edge. If the two planes are parallel, and either adjacent or offset, then their edge is reliable as a brightness measure, because the offset does not significantly change the distance brightness. However, if they are not parallel, then the new normal will give a different, incorrect brightness even though both edges are still constant. So a constant edge under one source is not particularly good. However, if two different sources are on the edge, then the change of normal will cause the ratio of the sources to change. So, when the diamonds are aligned by their transform, the flux points representing the two sides of the edge are not on the same line through the origin, so no brightness value could make them equal. Therefore, a constant edge under two or more sources, which aligns in flux space, is a reliable indicator of brightness. As with any other perceptual algorithm, coincidental alignment is possible. In two sources the locus of surface normals which happen to align form a plane through the sphere of normals, perpendicular to the principle arc.

An edge which is not constant, but has a constant ratio across the edge, is more common. Often two colour regions are part of the same surface, so that there is only a change of colour, but no significant occlusion depth or abrupt change of normal. There is probably some change of brightness along the edge, due to shadow, change of normal or distance, but the important thing is that the change affects both sides equally. This type of edge can also include cases where thin objects such as leaves or sheets of paper overlap each other. The edges produced by these cases are easily recognised and quite

reliable.

If the light on both sides of the edge is the same then, when the
diamonds are aligned by their transform, the edge values may be
equated by some brightness constant, as in the previous case of a
constant edge. However, in this case the flux is not constant, and
the fact that the changes on both surfaces trace the same line in the
flux structure is very good evidence that the lighting is common, and
so the brightness ratio is unity. This will even work reasonably
well in one source scenes.

A colour edge which occurs on a smoothly curving surface would
also be detectable as a constant ratio edge, provided that the
spatial resolution of the imager was good enough. It can also be
detected if the flux change near the edge is extrapolated.

The constant ratio test accurately selects which edges have unity
brightness. With it regions will be grouped if there is some path of
reliable edges between them which allows their brightnesses to be
compared. If all of the regions form one group, then there is
complete information to form a colour cube and complete the colour
and brightness constancy algorithms. This would happen in Mondrian
and undulating Mondrian worlds, and in some others. More typically,
the regions would form several disjoint groups, and each will be
surrounded by occluding edges or ones with sharp surface changes. If
a group was diverse enough, then it could form a colour cube of its
own, completing constancy in that area. Likewise, if the group
contains an area of mutual reflectance, then that corner may define
the brightness for the group. If two surfaces are thought to be of
the same material and so have the same brightness value, they could
also unify two groups like an edge.

- 157 -

When regions are gathered into groups with known brightness ratios, they may provide greater information than the regions separately. They will more completely fill a flux structure, and so better define an envelope, whether an envelope of normals or a diamond-shaped envelope. They have longer edges, to compare with neighbouring groups, and they provide a larger area when looking for distance gradients.

If processing ends with separate groups, and any is too small to form a colour cube on its own and also lacks a usable case of mutual reflectance, then it may be necessary to invoke spatial reasoning to determine the brightness ratio between the groups. Any method which will tell the relative brightness of either the reflectances or the illumination of two points from different groups is sufficient to unify the two groups, and so may be enough to complete the constancy.

In a practical implementation, the notion of critical regions and critical bands of brightness may be useful for directing attention and processing effort. Whenever a group of regions are formed into a colour cube, only a few of those regions will be involved in scaling the cube. This is especially true of the virtual white scaling. These might be termed the critical regions. If their brightness is incorrect then the whole cube is incorrect. An incorrect brightness on a non-critical region will only alter the perception of that one region. Now as the brightness of a critical region is reduced, there is some point when that region ceases to be involved in the scaling and its role is taken by another region. At this point the colour cube can make a major change of shape. The brightness where this occurs is the critical brightness for the region. If it is difficult to measure brightnesses, it is worth concentrating effort on the

critical regions and on the edge or path between such regions and their potential successors.

There are three bands in the comparison between two regions. Imagine for the moment that the colour cube has only two vectors. If one is so great that it is larger than the other in every component, then it scales the cube. As its brightness decreases, there is a point where one of its components becomes equal to a component of the second vector. This is the end of the upper band. Geometrically, the shorter vector just touches the edge of the colour cube. As the vector continues to decrease, the two vectors jointly determine the colour cube, until the end of the middle band, when the first vector is equal in one component of the second and less in all the others. After this, the colour cube is scaled only by the second vector. Now simply knowing that the brightness of the vector is in the upper band is enough to know the shape of the colour cube. The same is true for the lower band. It is only in the middle band that the shape changes, depending on the brightness ratio of the regions, and so it may be called the critical band. If a region ratio is not in a critical band, then a rough approximation may be acceptable. However, within the critical band, a precise brightness ratio is wanted.

Chapter Six

A Look at Secondary Reflectance

Determining Illumination in an Inside Corner

The Effect of Background Light

Measuring a Veiling Luminance

In a flat scene it intuitively seems correct that if every surface were made half as reflective, while the illumination intensities were doubled, then the same flux values would result. Contrast this with Gilchrist's rooms experiment [1979], where an observer was given a pin hole view into a miniature room. In one case the entire room was painted white but dimly lit, and in another an identical room was painted grey and seen under a hidden bright light. Subjects immediately recognised the true surface brightnesses, even when the average flux from the grey room was much brighter than the average flux from the white room. The same result was obtained when comparing a white-lit blue room with a blue-lit white room.

Gilchrist's own explanation was to observe that in the grey room the flux varies more severely from pixel to pixel than in the white room, and that the eye may somehow interpret this strong spatial contrast to guess the surface reflectance.

In this chapter we will consider two intuitively reasonable models for reflectance constancy in these and more general scenes, and will sketch their computational extraction. The computations suggested will be only simplified examples to show that it can be done, rather than the final form of the algorithms. The two sources of information are secondary reflectance and background illumination. In both cases the principle is the same, that once the illumination

has bounced off a surface in the scene its flux is measurable. If it then becomes an illuminant of another surface, then it is a case of an unknown reflectance under a known light, which is easily computable.

As an example, do a mental experiment. Imagine two inside corners, each under similar lighting; the scenes differ only in the brightness of the light and the reflectivity of the surfaces. There will of course be some secondary reflectance or mutual reflectance; some of the light reflected off of one surface of the corner will strike the other surface and make it brighter. The brightness will be strongest near the intersection. Given that the surfaces in one scene are less reflective than those in the other, increase the illumination on that scene until the direct or primary flux from it is the same as from the other scene. Now in each scene the first surface is providing the same amount of illumination to the second surface. However, in one scene the reflectance of the second surface is higher, so more of the light will be reflected.

What has happened in this demonstration is that the light has bounced off two surfaces, thus allowing any change in reflectance to twice effect the flux of the secondary, while the illumination still has only its one direct effect. Another way to explain it is that the secondary flux is the product of illumination and reflectance like any other flux, but in this case the illumination is directly measurable as the primary flux of the opposite surface.

In providing a computational model for extracting the surface and illumination colours and strengths from secondary reflectance, it is useful to narrow the range of scenes, discussing primarily ideal conditions. We will assume that all surfaces are Lambertian and

without texture; they receive and transmit light equally in every direction. We will also mainly consider secondary reflection between two intersecting surfaces, which thus form an inside corner, although useful information can be gathered from secondary reflection between non-adjacent surfaces. The typical corner to be examined is two intersecting planes where both the light source and the point of observation are distant points within the included angle. This situation avoids problems with shadows, perspective, and distance-from-source gradients. The angle may be acute, right or obtuse.

Let us call the reflectance vectors of the two surfaces R and S, the illumination colour I, with brightnesses i1 and i2 on the two surfaces, and let t1, t2, t3, t4 ... represent the transfer functions from one plane to a point on the other. We will look at the transfer functions shortly. Consider the components of the total flux, on each of the two surfaces.

reflectance	R	S
primary	i1 IR	i2 IS
secondary	i2 IS t1 R	i1 IR t2 S
tertiary	i1 IR t2 S t3 R	i2 IS t1 R t4 S
quaternary	i2 IS t1 R t4 S t5 R	i1 IR t2 S t3 R t6 S

This is an infinite sequence of bounces, but because the transfer functions t and reflectances R and S are all substantially less than 1 and because the transfer functions get weaker away from the corners as will be seen, the later terms become negligibly small. Notice that the even order terms have the same colour, so that the secondaries differ only by the scalar ratio i2t1 : i1t2. The odd order terms, such as the primary and tertiary have the same colour ratio, R : S, as the two surfaces. The odd terms also have similar

- 162 -

transfers, so that in the tertiary each has bounced both from R to S
and from S to R.

Using colour techniques it is easy to measure the brightness of
each bounce, or component of the sequence, since each has a different
colour. As an aside, it may also be possible to separate these
effects in a monochrome scene by measuring the flux of the region at
a point well away from the corner to get the primary, and then
subtracting this to get the secondary. This would assume constant
primary and negligible tertiary.

Having separated the components, it is possible to immediately
compute the colours, as separate from brightnesses, within the scene.
First, divide any intensity of the secondary vector by a typical
primary vector from the same region.

   i2ISt1R / i1IR = $(i2t1/i1)$ S

This gives the reflective colour of the opposite surface, times a
composite scalar.

In the present context this gives us once again colour constancy
without brightness constancy, although here it is done locally,
needing only the two regions. However, the same relation may be
useful in a different situation. If the surface colours have been
computed by other means, but the physical layout is not well known
then one may identify and measure the effect of a secondary without
knowing its corresponding region. The above computation gives the
colour of the region, and so restricts the source to only regions of
that colour.

Having computed the colours of the two surfaces, it is easy to
compute the colour of the illumination. For example, divide the
product of the two primaries by the colour of either of the

secondaries, which we recall are both the same.

$$(i1IR \; i2IS) \; / \; i2ISt1R \; = \; (i1/t1) \; I$$

This gives the colour of the illumination times a composite scalar. More precisely it gives the illumination striking the region in which the secondary was measured, i1 I, diminished by the first transfer function from the other region, t1.

Determining the brightnesses of the illumination and reflectivities is not as easy as determining the colours. If the three dimensional physical layout of the scene is immediately found, then it is straight forward to compute the transfer functions. Thus given t1, I, and (i1/t1)I, i1 is known, and using the primaries, the strength of R and S are known. In our mental experiment, it is likely that we could identify the physical layout and compare the corners with our previous experience of real world scenes. But it is also possible, and in this case easier, to compare the two corners with each other. This comes down to a similar problem that Land found. He demonstrated that colour constancy would be easy if it were possible to place a white card in the scene next to each object, and so compare known and unknown reflectivity under the same illumination. In our situation, we would ideally like to place a white corner having the same angle and orientation in the scene, and compare the transfer functions. However, in a different sense each of the two regions is a surface of known colour held up against the other. Since the angle of the corner determines the transfer functions, it may be possible to compute this angle by measuring several data points of the secondary component as it increases into the corner. We will now do this, using several simplifying techniques.

First, it should be said that the transfer functions which produce the different terms, t1, t3, t5, and so on, are actually the same. Each is a particular series of integrals that weigh and sum the light from all the points of the other region onto this particular point. The reason that we have named them differently is that the input to the first transfer is a constant function, and so its integrals can be greatly simplified. For each of the other transfers, the input is the non-constant brightness function produced by the previous transfer. Each function is in fact $t(p)$, with a value for every point p in the region.

We now describe a simplified view of the transfer functions. Consider first the light given off by a single face. Being Lambertian, the brightness of the surface per viewing angle does not change with either the distance from the surface or the angle of view relative to the surface. This means that the amount of light received from the surface is solely a function of the apparent viewing angle. Let us imagine that the surface is much longer than wide, in order to see how that width affects viewing angle in two dimensions. It is well known that for any two points and some angle, the locus of all points which are the apex of that angle with those two points describes part of a circle which meets those two points. In the case of a right angle, the circle described is the hemisphere above the diagonal described by the two points. If we now take the cross section of the elongated surface, so that the top and bottom edges are the two points, then the iso-values of the illuminating strength can be seen. These are circular arcs from top to bottom that have their centre on the perpendicular bisector.

In general, if the amount of visual arc of one surface is seen

from another, and the first surface has a constant and visible flux, then the amount of illumination on that point of the second surface is easy to determine. It is flux seen in a known camera arc, scaled up by the ratio of surface arc to camera arc.

We now consider how the second surface intersects this
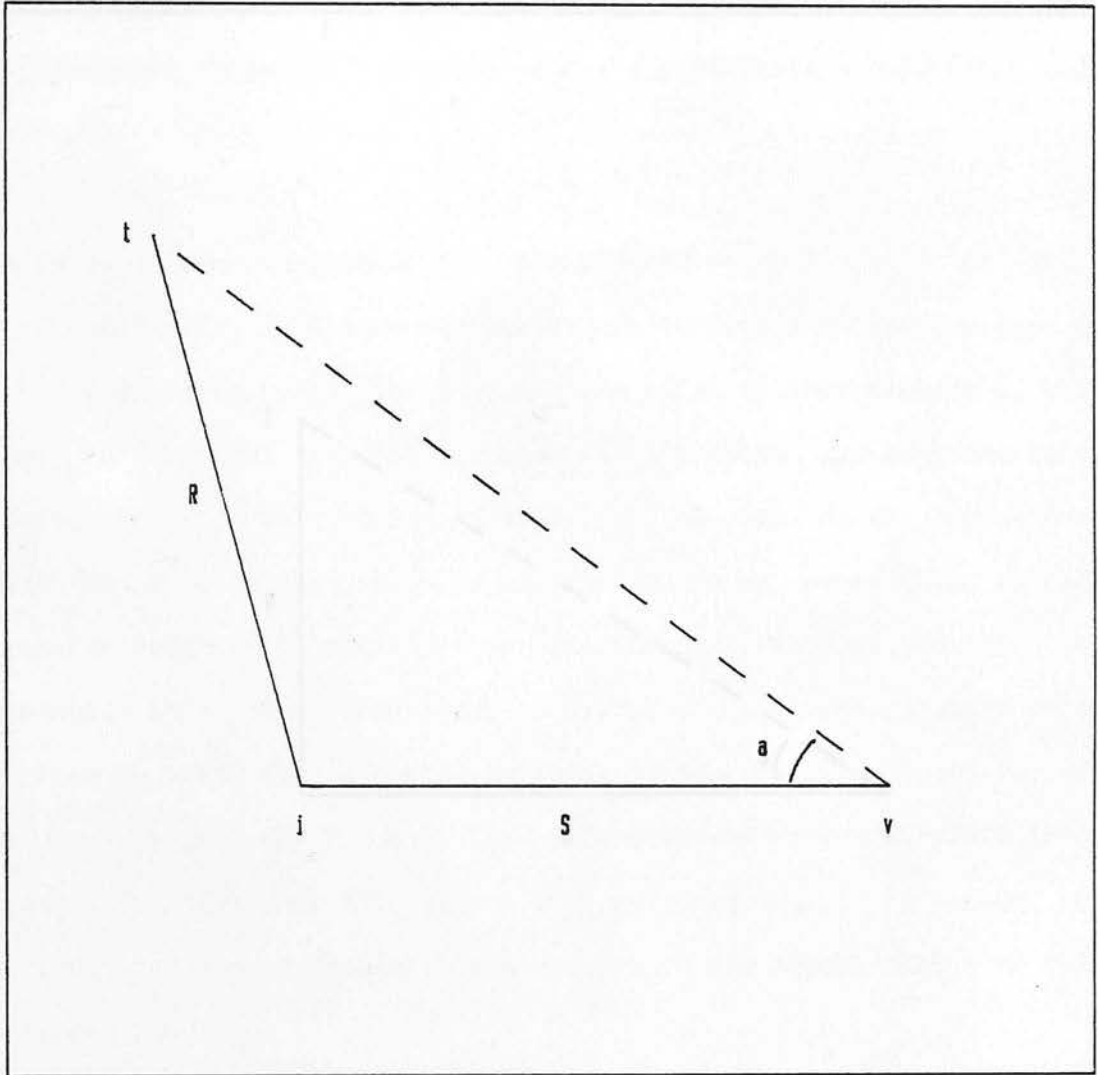
6.1 Profile of an Inside Corner

illumination function from the first surface, by looking at the cross section of the angle, taken normal to the centre of the intersection. From the diagram 6.1 it can be seen that as point v moves along surface S, the amount of light received from surface R is a function

of the angle to the furthest point, t, of R. Since only the top edge, and not the angle or point of intersection, is affecting the brightness on point v, then the rest of R can be ignored for the moment, and the perpendicular P, to S through t can be drawn. Its length will be called p. This is shown in diagram 6.2. Now the
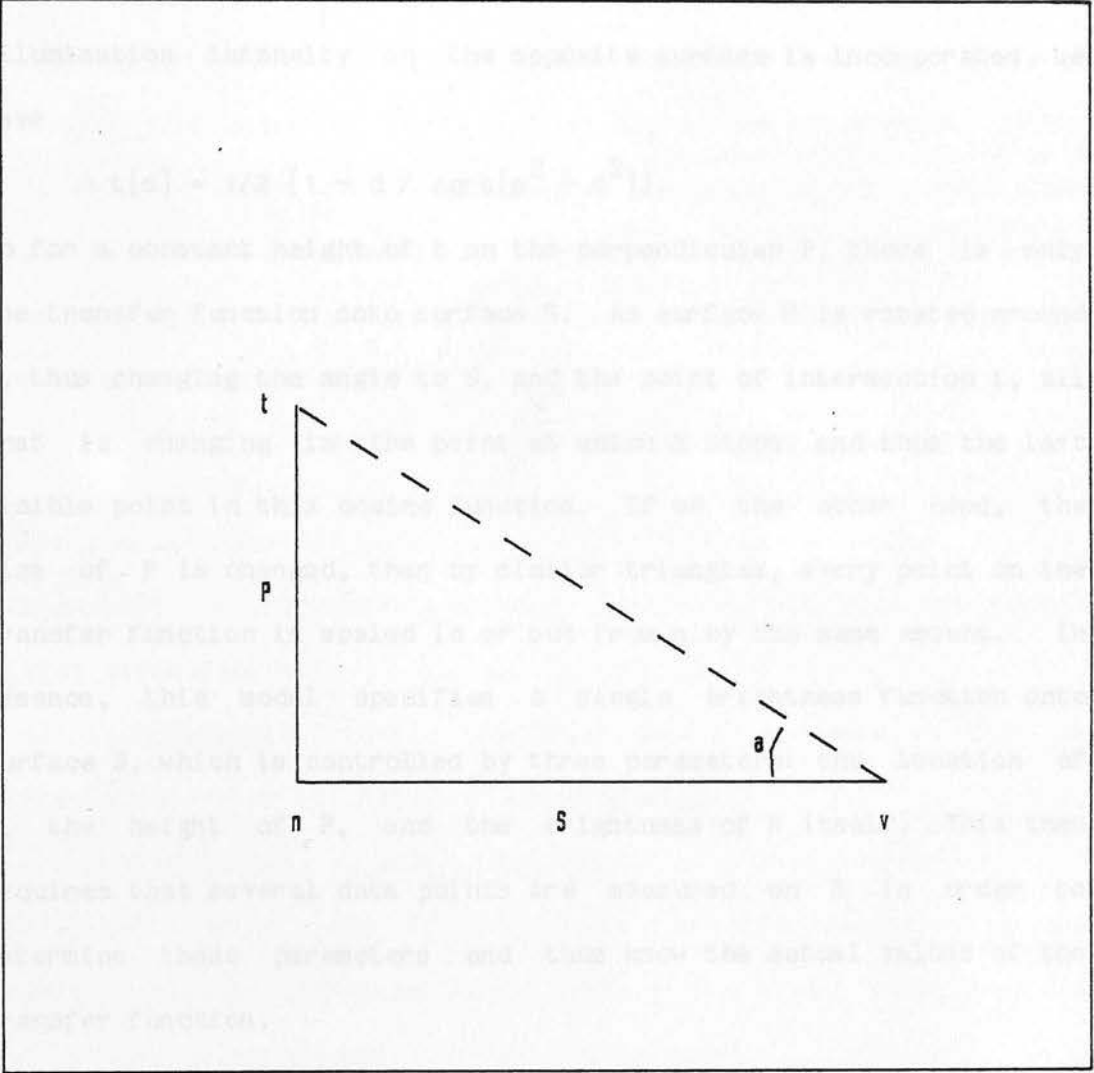
6.2 Mutual Reflectance is Equivalent to some Perpendicular

visible arc from point v as it moves along S can vary from 0 degrees at the extreme right to 180 degrees in the extreme left. For this same range to be expressed as the proportion of the available view occupied by R, that is in a range from 0 to 1, then the function

would be

$$\frac{1}{2}(1 - \cos a),$$

and the cosine is a function of the ordinal p, and the abscissa v-n which we will call d; in particular

$$\cos a = d \, / \, \sqrt{p^2 + d^2}.$$

When these two expressions are combined, and the value i of the illumination intensity on the opposite surface is incorporated, we have

$$t(d) = i/2 \left(1 - d \, / \, \sqrt{p^2 + d^2}\right).$$

So for a constant height of t on the perpendicular P, there is only one transfer function onto surface S. As surface R is rotated around t, thus changing the angle to S, and the point of intersection i, all that is changing is the point at which S stops, and thus the last visible point in this cosine function. If on the other hand, the size of P is changed, then by similar triangles, every point on the transfer function is scaled in or out from n by the same amount. In essence, this model specifies a single brightness function onto surface S, which is controlled by three parameters: the location of n, the height of P, and the brightness of R itself. This then requires that several data points are measured on S in order to determine these parameters and thus know the actual values of the transfer function.

However, without finding the transfer function it is possible to find out the relative illumination strength, i1 : i2, between the two surfaces. This computation requires less assumptions about the scene. Consider what happens when the point v closely approaches the corner. At the limit the height of the opposite surface becomes irrelevant. Here the zenith arc is divided into two parts, one that

is the size of the angle, and is open to primary illumination, and the other that is the complement, and receives secondary illumination from the opposite surface. Therefore the amount of secondary is only a function of the angle. If indeed the surfaces are planes, then the troublesome third dimensional effects also vanish, since it approaches the ideal 180 degrees.

Having established the limit of the secondary on this surface, consider the opposite surface as it approaches the corner. It too will be governed at the limit solely by the angle at the corner, but this is one and the same angle! What has happened is that t1 and t2 have converged, so the ratio of the secondaries becomes the ratio of the illumination strengths.

$$i2ISt1R : i1IRt2S \qquad as \ t1,t2 \dashrightarrow t$$
$$= i2IStR : i1IRtS$$
$$= i2 : i1$$

Notice that the secondary is stronger on the surface with the weaker direct illumination.

It is now possible to find the brightness ratio between the two reflectances. For the ratio of the primaries is

$$i1IR : i2IS = i1R : i2S$$

and we now know i1 : i2. If we now combine this result with the previous result where the colour ratio of R : S was found, then the full reflectance ratio of the surfaces is known.

Indeed, all that is not easily obtainable is the absolute brightnesses of illumination and reflectivity. The missing result is the trade off between a scene of bright illumination with dull reflectances and one of weak illumination and bright surfaces. This is surprising because that is one of the pieces of information that a human would seem to pick out most easily.

The ratio of the illumination brightnesses has another use. It identifies the angle of the light source relative to the two surfaces. It is well known that illumination brightness is a function of the angle of the surface to source. In particular, the brightness is diminished by the cosine of the angle between the surface normal and the source. The ratio of these cosines from the two surface normals must be the same as the ratio i1 : i2.

We now turn to look at background illumination. In a sense this is the opposite extreme form of secondary reflectance. Mutual reflectance, as in the corner above, comes from one region and takes on that particular colour; it seems almost "focussed". In contrast, ideal background has a global brightness and colour, which it takes on everywhere, and which comes from no particular region. One example is sky light, where dispersion of sunlight in the air can cause almost even illumination from the hemisphere above the horizon. A room could also approximate even background light to the extent that portions of it receive similar primary illumination and have similar average reflectances. For our purposes, we will take ideal background illumination as a given, in order to see what use it might be.

First consider the effect of background light on the flux cube. If the background has a constant colour but a variable strength, then it should be regarded as another illuminant, and the flux cube processing done with one more dimension. If the colour is not constant then it is not background at all, but a set of different coloured sources, which may for example be treated as cases of mutual reflectance, or as noise. It would also be possible to treat them as local sources, where each influences only a small part of the plane.

However, our present case is constant colour and brightness, and this would be added equally to every point in the scene. This will shift the flux plane sideways by a fixed amount giving a new parallel plane. Assuming that the colour of the background is independent of the colours of the other sources, this means that none of the flux planes now will pass through the origin. The normal to the plane will remain constant.

To begin to get an intuition for the behaviour of this shift, consider two regions under the same sources and an independent background. The regions are of identical reflectance colour but one twice as reflective as the other. The resulting diamonds will be parallel, but the brighter one will be twice as large, and its bottom corner will be twice as far away from the origin. It is obvious that in the case of parallel planes, relative distance from the plane to the origin is the same as relative brightness.

However, as the reflectance changes colour, so does the direction of the background vector, since the vector is the background illumination times reflectance. When a plane is offset, we do not immediately know the measure of the background vector, since any vector from the origin to any point on the plane would act as an offset vector to generate that plane. For a given plane, the true background vector can be determined or partially determined in several ways. First of all, the background vector can not be greater than any of the flux points in the plane, thus providing an upper bound. Obviously the background vector must be non-negative in each component, which is its lower bound.

A second technique is to identify the flux diamond. In the ordinary flux cube it is known that the right and left edges of each

diamond meet at the origin. Therefore it is only necessary to take the extreme points in each direction in order to describe a line; the second point is provided by the origin. Once the diamond has been shifted, two points are needed to define the edge, and preferably more for confidence. This is easiest when some points are identified as being in shadow relative to the other source because shadow pixels are always on the edge of the diamond. If they vary in flux then they provide multiple flux points and so they will define the diamond edge. If this occurs for both sources, in two different portions of the region, then the defined edges triangulate the location of the shifted origin. Of course, when both sources are in shadow in the same portion of the region, then the background alone can be directly measured, and this is the tip of the diamond and the value of the background vector.

If the location of the origin is known on one plane, then the origin on all the other planes is partially known. In any plane parallel to the one of known origin, the diamond origins are all co-linear with the known one. In non-parallel planes, the origin must correspond to the known origin under the transformation between the two planes. This restricts the origins to a computable line passing through the point where the plane comes closest to the cube origin. It would only require one of the diamond edges to then triangulate the diamond origin in the new plane.

In either of these techniques, the actual background vector for that region is found. When the background vectors for two regions are found, the component-wise ratio of the two vectors is the transformation between the regions; it is both the colour and magnitude of the reflectivity ratio between them. As shown

elsewhere, the magnitude information can be added to the colour cube to refine it. If all the relative magnitudes are known, then the uncertainty is removed from the cube and it gives better absolute colour and brightness information. Indeed, the background vectors themselves may be used in a colour cube. Since they do not have an uncontrolled brightness scalar, then the naive cube result is correct.

A third technique for using background offset in colour constancy is to compare the distances from the plane to the origin. We will now show that any transformation away from maximum reflectance will reduce the offset, independent of changes to the angle of the plane. This method takes as input the normal to the plane, and the least distance from the plane to the global origin. We begin by examining the effect of a reflectance transformation on the normal and the offset.

Let A and B be the source flux vectors, and G the background vector, all three independent. The normal to the plane, N, is the cross-product of A and B.

$$N = A \times B = (a2b3-a3b2, a3b1-a1b3, a1b2-a2b1)$$

The length of the vector produced when G is projected onto N is the distance from plane to origin at the normal, its closest point. The component of the projection is computed using the dot product as

$$c = (G.N)/(N.N)$$
$$= (g1*n1 + g2*n2 + g3*n3) / (n1^2 + n2^2 + n3^2)$$

so that the vector cN is the shortest vector from the origin to the plane. Its length gives the distance, and is computed as

$$length(cN) = sqrt(cN.cN) = c \; sqrt(N.N) = c \; length(N).$$

Let us consider this same illumination on another region in the scene, that is after a transformation T to another reflectance

vector.   Each of A, B and G become TA, TB and TG, respectively.   The
new normal is called N' and the new component c'.

$$N' = TA \times TB$$
$$= (t_2t_3(a_2b_3-a_3b_2), \; t_1t_3(a_3b_1-a_1b_3), \; t_1t_2(a_1b_2-a_2b_1))$$
$$= (t_2t_3 * n_1, \; t_1t_3 * n_2, \; t_1t_2 * n_3)$$

$$c' = (TG.N')/(N'.N')$$
$$= (t_1g_1 * n_1' + t_2g_2 * n_2' + t_3g_3 * n_3')$$
$$/ (n_1'^2 + n_2'^2 + n_3'^2)$$
$$= t_1t_2t_3(g_1 * n_1 + g_2 * n_2 + g_3 * n_3)$$
$$/ ((t_2t_3n_1)^2 + (t_1t_3n_2)^2 + (t_1t_2n_3)^2)$$
$$= (g_1 * n_1 + g_2 * n_2 + g_3 * n_3)$$
$$/ (n_1^2 t_2t_3/t_1 + n_2^2 t_1t_3/t_2 + n_3^2 t_1t_2/t_3)$$

This gives c the same numerator as c'.

We wish to demonstrate that $length(cN) > length(c'N')$ when
$t_1, t_2, t_3 \leq 1$, and strict inequality holds for some t.   The desired
result

$$sqrt(cN.cN) > sqrt(c'N'.c'N')$$

is true exactly when

$$cN.cN > c'N'.c'N'$$

For simplicity we use the form of c' which has the same numerator as
c.   This numerator appears squared on both sides of the relation, and
so being positive, it may be divided out, leaving us to prove

$$(n_1^2 + n_2^2 + n_3^2) \; / \; (n_1^2 + n_2^2 + n_3^2)^2$$
$$>$$
$$((t_2t_3n_1)^2 + (t_1t_3n_2)^2 + (t_1t_2n_3)^2)$$
$$/ (n_1^2 t_2t_3/t_1 + n_2^2 t_1t_3/t_2 + n_3^2 t_1t_2/t_3)^2$$

Cancelling in the first expression, and multiplying by   the
denominators, which are both positive leaves us to prove

$$(n_1^2 t_2t_3/t_1 + n_2^2 t_1t_3/t_2 + n_3^2 t_1t_2/t_3)^2$$
$$>$$
$$(n_1^2 + n_2^2 + n_3^2) * ((t_2t_3n_1)^2 + (t_1t_3n_2)^2 + (t_1t_2n_3)^2)$$

- 174 -

Expanding terms, reducing and re-ordering provides

$$n1^4 *t2^2 t3^2 /t1^2 + n1^2 n2^2 *t3^2 + n1^2 n3^2 *t2^2$$
$$+ n1^2 n2^2 *t3^2 + n2^4 *t1^2 t3^2 /t2^2 + n2^2 n3^2 *t1^2$$
$$+ n1^2 n3^2 *t2^2 + n2^2 n3^2 *t1^2 + n3^4 *t1^2 t2^2 /t3^2$$

$$>$$

$$n1^4 *t2^2 t3^2 + n1^2 n2^2 *t1^2 t3^2 + n1^2 n3^2 *t1^2 t2^2$$
$$+ n1^2 n2^2 *t2^2 t3^2 + n2^4 *t1^2 t3^2 + n2^2 n3^2 *t1^2 t2^2$$
$$+ n1^2 n3^2 *t2^2 t3^2 + n2^2 n3^2 *t1^2 t3^2 + n3^4 *t1^2 t2^2$$

It will be seen that the inequality holds for each of the nine corresponding pairs of terms. The first three differ by t1 squared, the second three by t2 squared, and the third three by $t3$ squared. This completes the proof.

This result allows us to immediately estimate the colours of the illuminants, by finding the plane with the greatest offset from the origin. The plane through the origin parallel to this will be the closest to the plane defined by the two sources. It will differ to the extent that the surface it represents is not "pure" white.

It should be noted that this last technique can not be used when the background illumination is not independent of the colours of the sources. In this case, all the planes will still pass through the origin, and so there would be no perpendicular offset to measure. However, there would still be a shift within the plane, which could be measured using the previous techniques.

These techniques for using the background shift have been described from the point of view of extracting colour constancy information directly out of raw flux data without computing a colour cube. However, the colour constancy techniques described earlier are more reliable. Perhaps the best use is in combination where the main techniques perform the colour constancy, while the background shift techniques explain and compensate for the effects of more realistic

- 175 -

background and secondary reflectance.

When it was originally developed, the method of using the background shift for directly determining reflectance was simply of interest as a mathematical curiosity. Many of the early studies in computational vision, in particular many blocks world models explicitly assumed just such a perfect ·constant background illumination. It is interesting to show that that assumption alone can be sufficient for determining surface reflectances, even though the small size of the background shift would have to be enhanced by very good imaging hardware, a bright background, or clever error reduction. Therefore, it was surprising to realise that the background shift, with slight modification, accounts for Gilchrist's results [1983b] on veiling luminances, discussed in chapter two. Recall that a strong light may be added to every point in a three dimensional scene without altering the correct reflectance perceptions, even though all of the edge ratios are changed dramatically. However, if a Mondrian is used, then the perceptions change in accordance with the new edge ratios.

To gain insight into this result, consider the appearance of the images in flux space. In this particular experiment, each Mondrian region seems to have nearly constant flux, and so maps into approximately a single point in flux space. Using the rule of minimum light sources and the fact that each point is non-dimensional, it is assumed that there is a single colour of illumination and that each point corresponds to a separate region. There is no evidence to indicate that these points are not simply on lines through the origin, as with any other usual flux structure, so this assumption is made, leading to incorrect perception of the

- 176 -

reflectances. The only unusual property of the flux space of this image and that of any other flatly lit Mondrian is that here no flux point falls in the lower half of the space in any of the three dimensions. This also is true of any Mondrian that is composed entirely of bright surfaces, which is in fact how subjects perceived Gilchrist's veiled Mondrian. It is possible for a constancy algorithm to detect this property and correctly adjust for it, but this would mean incorrect perception of each Mondrian-like image that does not contain quite dark surfaces. The experimental results show that the eye does not choose that option.

In contrast, the flux space of the real scene contains complex flux structures, because of the large changes in brightnesses within a region. The structures in the outdoor scene will be mainly planar because of the two colours of light source - sunlight and skylight. The structures of the indoor scene will be linear because of the single source. However, an observer familiar with representations of flux space would notice something very odd. All of the flux structures would converge at a single point, but that point would be far away from the origin of the flux space. This shift of the origin is only consistent with a universal increment of flux. It can not be explained as an un-naturally large background shift, partially because of the direction of the shift, but also because the shift is constant for each pixel, whereas background shift is dependent upon the reflectance.

There are at least two different measurable changes in the flux space caused by the veiling luminance, the shift of the structure convergence point, and the shift of the shadow convergence point. The first has already been mentioned. This shift can be observed for

- 177 -

even a single region, because the flux structure defined by the illumination gradients on the region define a line or plane which does not pass through the origin. This suggests that a shift has occurred, but does not fully find the new convergence point, which is only constrained to lie on the line or plane of that structure. The addition of another region, which will have a different and intersecting structure, actually measures the shift. For any given shift, if the original scene has only one illuminant colour, then there is only one reflectance colour which will remain co-linear with the origin, and all others will show a shift. With multi-dimensional illumination, the reflectance colour is not necessarily unique, but in a negligibly small set, and there is more shift information available from the transforms and the shadow convergence, discussed below.

It may sometimes be useful to regard the origin shift as having two components, one in brightness and one in colour. This is particularly relevant to Gilchrist's experiments, because it was assumed that the tungsten veiling luminance would be equivalent to the sunlight effects on the scene, since both are called white light. The actual expected shift would be primarily in the brightness dimension, but also offset slightly due to the difference in veiling colour. It is not clear whether the eye only compensates for the brightness component, or whether, as expected by the ecology of the problem, it compensates for the shift of origin independent of its colour.

The second type of shift that can be detected is the shadow shift. When there are multiple source colours, the occlusion of a source will cause the flux points to lie on the lower edge of the diamond.

This edge must pass through the origin. Therefore, when shadows are detected and the shadow edge is not constant, there are at least two different points on the edge, and this is sufficient to define the line of the edge. Such shadow edges constrain the location of the shifted origin in a similar manner to the flux structures themselves. Notice that shadow edges shifted by a veiling luminance are even less likely to be co-linear with the origin than are the flux structures themselves. If both shadow edges appear on a single region, then the shift can be measured from that region alone.

Practical computation of the shift should not rely on a single region, but should collect all the evidence in order to best locate the new origin. As usual, the results should not be blindly averaged, but should be compared to be certain that they are consistent, and that they agree on a single origin. Then the averaging should take into account the reliability of each bit of evidence, since for example larger flux structures more accurately define a line or plane.

So, by examining the flux structures produced by veiling luminance on both Mondrian and real world scenes, it is no longer a surprise that it is possible to discount the veil in only the more complex scenes. The real surprise is that the human eye makes use of this information, since veiling luminances are rare in the natural world, perhaps confined to the surface of water on cloudy days. However, Gilchrist has suggested a good reason for discounting the veil, which is intraoccular scatter. This is the substantial amount of light which is reflected about within the eyeball, and which tends to evenly illuminate the retina. Some commentators have suggested that this scatter can be used to compute constancy, since the scattered

light represents the average colour and brightness of the image. However, it is clear that the scattered light is then just a quick way to find the average flux and we have seen in chapter two that humans do not compute constancy from the average flux. In addition, it is a poor constancy measure because scene reflectances do not necessarily average to white and because the illumination may vary greatly within a scene, especially when examining a detail in a shadowed portion of the scene. So, intraoccular scatter is likely to distract from reflectance constancy rather than enhance it, and this may be why the eye is able to correctly process veiling luminances, given sufficient image information to detect the veil. This hypothesis would be supported by experiments that combine and compare both veils and directly caused intraocular scatter. For example, are the same Munsell matches obtained from the Mondrian when light is projected into the eyeball peripherally?

There is a second ecological reason for processing veils. A common problem with imaging systems used in computational vision is that the slope of the response curve is such that a zero flux value does not necessarily correspond to black. The common solution is to take the differences between flux values to cancel any black bias. However, this shift affects the flux structure identically to a veil, except that black bias can also occur in a negative direction. It could easily be the case that the eye is susceptible to black bias, and like an electro-mechanical system it could vary with age and other factors. Vision theorists since Wallach have assumed that any such problem would be avoided because the fluxes are not important, only the ratios or comparisons of fluxes. However, Gilchrist's results have shown that ratios and comparisons as previously defined

are not enough. It would be most interesting to repeat Gilchrist's results using a veiling luminance of negative strength, to see if that is also correctly processed the same. This presents a technical problem, since most methods of reducing the light are not constant but flux dependent. Thus neutral density filters subtract a percentage of the local flux rather than an equal amount at each point. A possible solution is to use photographic processing of an image of the scene.

Although such experiments have not been done in the present research, interesting results were obtained from presenting veiled data to the computer implementation of flux space discussed earlier. As expected, a veiled Mondrian scene produced the same results as human subjects, with the perception incorrectly matching the ratios rather than reflectance. A veil over a complex scene with two illuminant colours was correctly processed to discount the veil, whereas a similar scene with only one illuminant totally failed to produce a meaningful image. After some experimentation, the reason was determined. The subroutine which processed linear flux structures, which was one of the first built, had an implicit assumption that all such structures pass through the origin. Thus almost every distinct flux value in the scene was considered to be a different reflectance value. In contrast, both the routine which determines the dimension of the illumination and the one which fits planes to flux data did not have a hidden requirement that each structure pass through the origin, because of the way they find a basis for a set of flux points. Therefore these later routines processed veiling luminances without difficulty, but likewise without even noting that the veils existed. This in fact prompts a third

explanation of why the eye can process veils, which is simply that it may be a fortuitous artifact of the implementation of constancy. More exactly, since constancy can be done either with or without the requirement that the flux structures pass through absolute black, the eye may well have evolved without that requirement, which may well be the computationally simpler method in the context of biological vision. Needless to say, the linear-fit subroutine was modified to only force a structure through zero in the absence of any other information, and the correct perception was obtained.

For completeness, the alternative treatment of Mondrian-type ambiguities that was suggested above was also tested. In this case, if none of the structures is linear, then the origin is shifted to the black hull, which is the largest point which is not greater than any measured flux. Any larger point would imply that some surface reflects by a negative amount. This modification did not provide particularly good results. The first scene tested had no reflectance value of less than twenty percent, so every object other than white was perceived as darker than its correct reflectance. Also, there was a colour bias because the minimum reflectances among the channels were not the same. Both effects were more pronounced in darker objects. It must be noted however, that the results were always less in error than those achieved without any origin shift because the magnitude of the veil was greater than the un-blackness of the least reflective objects. Of course, when a black object was added to the scene, the results were correct.

Chapter Seven

Additional Topics
Proposed Future Experiments
Context

This chapter begins with a discussion of a few remaining topics. It will be shown that although the flux space techniques can depend upon small differences of flux, they also allow powerful error correction and smoothing techniques, and so might be implementable on currently available imaging systems. Mixed pixels occur in every image when a scene edge falls across a pixel. These are often considered troublesome, but colour methods allow them to be easily recognised, and then they may be used to estimate the precise subpixel location of the edge. Within the flux structures, the colour angle between sources varies, but the changes are only useful for determining the available resolution of flux.

Although it is probably possible to continue extending flux space theory, there are two other areas of development which should be allowed to catch up, so that there will be an exchange of results. These are perceptual experiments and implemented algorithms. Much of the central section of this chapter is devoted to proposing experiments related to flux theory. Toward the end of the chapter there are a few suggestions on how to go about building a flux space program.

The experiments tend toward three purposes. The first is to see how much of flux space and envelope information is extracted by human perception. The second is to refine the experiments of Land and Gilchrist in order to better define the capabilities of human colour

constancy. The third is to refine our understanding of the capabilities of human brightness constancy, in order perhaps to find clues to possible mechanisms for this missing link. Some are presented as detailed procedures with explicit goals, while others are vague notions that a given area may be worth studying. At the beginning of the experiments there is a description of how to build scenes to isolate any combination of brightness gradients, or even how to have different sources exhibit different types of gradient within the same scene.

After the experiments, flux space theory is restated according to the conceptual steps involved in its development, rather than according to its results and capabilities.

Initially it seemed that flux space techniques would be particularly vulnerable to signal noise and to errors in the flux values received. However, we will now see that it also provides unusually strong error correction facilities. Here we will be discussing mainly errors which are independent in each channel and at each pixel. These errors may be statistical noise, or occasionally may bear no relation to the signal at all. The techniques do not all apply to errors which systematically affect all the colours at a pixel. However, these errors can be corrected late in the algorithm because they will appear as single pixel regions with a different colour from their surround. None of the corrections applies to scenes with systematic, rather than independent, errors over multiple pixels.

At the most primitive level, the colour planes can be smoothed simultaneously. If most of the channels are proportional to their neighbours, then any channel which is not is likely to be wrong can

be smoothed into line.

To this can be added knowledge of edges. It is easy to find the regions, and it is meaningless to smooth across an edge, so the smoothing mask can be confined to single regions. However, this still leaves many problems. For example, shadow edges will still be smoothed, and strong gradients or cross gradients will be incorrectly smoothed unless more intelligent masks are used.

Many such ad hoc techniques were developed early in the present study, using the contemporary obsession with masks, and prior to the discovery of flux space. From flux space theory we know that all of the points in the region should lie exactly on the flux structure, which is a plane in the case of two sources. The structure is further restricted to passing through the origin, which is helpful in reducing noise in the darker regions, where it is usually most significant. Therefore, the first approach to noise reduction using flux space is to move each errant point back into the plane.

However, a much better alternative can be suggested. Consider an error in one channel only. The displacement out of the plane will be orthogonal to the flux dimensions, but will not be perpendicular to the plane. Therefore, correcting it to equal the nearest point in the plane will always be wrong. Fortunately, there is enough information in flux plane theory to determine which of the channels are in error and by how much. This involves adding spatial information. Within each separate illumination colour, the flux will usually be a determinable function around the pixel. By interpolating in each source colour and then summing the sources, a close estimate of the correct flux is obtained. Notice that this can improve a pixel where every colour is wildly in error, such as might

happen with a CCD camera where single pixels sometimes cease to correlate to flux received. A small amount of computation can be saved by interpolating on the channels themselves, but this can increase the error around shadow edges and strong gradients.

Clearly it is better to do region-wide correction rather than correction over small masks because there are more cells contributing to the average. This of course assumes that the average of the measured flux is the same as the average of the true flux. In some cases it should be possible to do scene-wide correction. Certainly when there are three or more distant point sources, every pixel of the scene is constrained to lie on the same envelope. The approximation to this envelope is best taken over the entire scene, and so the errors can also be reduced scene-wide.

Related to error correction is the problem of quantisation. The flux values from mechanical imagers are normally received as digital quantities, usually with 256 levels. Clearly there is substantial rounding of values. However, for some purposes this may be regarded as simply a form of noise; the algorithm can do region or scene-wide fit to these levels and determine more precise floating-point values. Of course, it would be better to go directly from the analogue flux to the floating-point values.

Visual receptors always look at a small area of scene rather than at a single point. If an edge occurs in that area then it produces a mixed pixel, which is an interpolation of the two different fluxes. These are unavoidable. There are several other related practical problems which are of less interest here. With a videcon camera, a strong edge can alter a few subsequently scanned cells by bounce or overshoot. Using any camera type, if the catchment areas overlap, it

- 186 -

is possible for a single sharp edge to generate a band of mixed pixels two or more wide. The following suggestion for processing mixed pixels assumes ideal behaviour, including linear receptivity across the catchment area, but it can easily be modified to a different, but known, camera response.

It is easy to approximate the value which the cell would have if one of the surfaces continued across the cell. One way is to use the value of its neighbour. Another is to extrapolate from the flux changes in the area. Either way, one value is obtained for each surface adjacent to the mixed pixel. There may be several bordering regions in the case of a corner, but only edges will be described here.

The two extrapolated fluxes can be found in flux space. Ideally, the mixed pixel will lie on the line segment between them. However, it may be off the segment because the catchments of the receptors in the different colours do not align precisely. In general, taking the distance of the flux in the direction of the line estimates the proportion of the two regions in the mixed pixel.

This proportion then approximates the location of the edge within the catchment area. If the amount of distance between the edges of adjacent catchments is known for the receptor system, then the location of the edge can be computed with substantially better than one pixel resolution. If resolution is especially important, then line smoothing along the edge may help, or information along the edge may be integrated at the proportion stage to provide intelligent noise reduction for the whole edge.

One of the many dead ends within flux space theory was the study of the width of the flux structures. In two-source scenes, this is

simply the angle between the two lower edges of the diamond. Since it changes with the region reflectivity, one might wish that, for instance, it might be widest when aligned with the axis of the colour cube. Unfortunately, this is not the case; the width of source angle is almost totally uninformative.

We state without proof the observation that for any scene in two sources, the angle is maximum as it approaches one of the axis faces of the flux space and is bisected by the forty five degree angle on that face. This is true for only one face, which happens to be related to the three classes of lighting seen in chapter three. The face is the zero locus of the colour component which has the middle component-wise ratio between the source colours.

In fact, the width is tied to the angle of the plane, which is already established and used in flux space theory, and so the width carries no more information. It would never occur that it is possible to find out the angle width without also knowing the slope of the plane. However, notice that as the angle collapses, the resolution of points in the flux structure deteriorates; studying the relationship between angle width and structure slope can allow the accuracy of flux space measurements to be predicted from the slope of the flux structure, and so from the relative surface colour.

Before suggesting specific experiments, it is useful to consider the range of scenes which may be presented to a colour constancy algorithm. An important deficiency in previous work has been the narrow range of scene types considered, and these are usually at the barren end of the scale.

One way to classify scene types is by what types of brightness changes may occur in the scene. Brightness may change because of

distance from source, shadows, or change of surface normal. Shadows
may be subdivided into focussed and unfocussed, according to whether
or not penumbra are generated. Change of normal can also generate a
special type of shadow called self-occlusion. We will now see that
it is possible to create a scene having any permutation of these
brightness gradients.   This is useful because it is thought that
different gradients may have different roles in brightness constancy.
If they are separated in different scenes, then the success of an
algorithm will give clues about its mechanisms.

Land's Mondrian world has only distance changes.   We will
characterise Mondrian worlds as having no change of normal, and so no
self occlusion. However, there are other Mondrian worlds beyond
Land's.   Shadows may be added by objects outside the scene, in line
with a source. These could be made with or without penumbrae.
Distance gradients may be removed by using a distant point source
perpendicular to the object plane.   So Mondrian worlds, that is
worlds without normals, can easily be made with any combination of
the other brightness changes.

A way to create a world with brightness gradients due to normals,
but no self occlusions or shadows is to have depth changes which are
less than the angle to the source. Such a world might be called an
"undulating Mondrian", but it is not a type of Mondrian world since
it has changes of normal. This is probably a very interesting world,
since it has no shadows, which are thought to be a strong simple clue
to the lighting.

It is possible to add spatial occlusions with respect to the
viewer without adding shadows by facing the occlusion towards the
source. Each source which is not perpendicular to the surface

eliminates occlusions from a 180 degree arc. Within the undulating Mondrian, distance gradients can be added or avoided as before according to choice of close or distant source. Likewise, it is possible to add focussed or unfocussed shadows without generating self-occlusions, by adding objects outside the scene as before. This completes the eight permutations of the three brightness gradients.

Notice that having self-occlusions without shadows is difficult. It is possible if the self-shading object runs off the edge of the scene, or if it makes a second change of normal to return to being lit, or if another object is closer to the camera and occludes the self-occluding object.

In addition to creating different combinations of causes of brightness changes with respect to the scene, the sources may be selected so that they display different brightness gradient properties. For example, in Land's Mondrian, one source colour could be constant while the other is side lighting that generates a distance gradient. This experiment asks whether the second, constant source can be detected, or whether the eye perceives only one source. If it is perceived as one source, then the constant light could be interrupted to generate a shadow, which is likely to change the perception.

In general it might be the case that colour constancy occurs more as a hierarchical rather than heterarchical process. If this is true then having say one light only exhibiting normal gradients while another exhibits only distance gradients might not be correctly perceived, because the modules which process these two types of changes are not interconnected. It may also be that some of the gradients are first processed through a spatial understanding

process, while others are not. If it is possible to make constancy fail in other than extremely degenerate cases, this will have useful implications.

The difficult gradient to separate with respect to the sources is the one caused by change of normal. In general, every change of normal is registered by a flux change in all the sources. Since this is such a universal phenomenon, it is likely that both colour constancy and spatial understanding make use of it. We suggest three contrived alignments where the source fluxes do not all change together with a change of normal. The first is to hide all normal changes in shadows; every region which is not planar is lit by only one source. It could be used to have only a single source change by normals, or to have several sources change, but each in an independent region where the other sources are in shadow. The second way to separate changes of normal is to have the normal constant with respect to one source, even though it changes in universal co-ordinates. This can be done by putting all the points on a sphere centered on the source to be made constant, or by using a cone which points directly toward that source. Although even harder to align, this method has the advantage that the areas are under more than one source, one of which is constant. A third method is to have the normal and brightness gradients cancel each other. It is a particularly difficult surface shape, which gives a constant flux from one source, but gives surprising normal gradients from any other source location.

Gilchrist was partially right about needing spatial information to process colours, in that spatial information seems to be needed in determining how much of the edge brightness ratio is due to

- 191 -

reflectance versus illumination. However, his implication is that spatial perception occurs first, followed by colour perception. This is clearly not always true. For example, flux space methods allow the number of sources to be determined immediately. This is more a by-product of the the colour processing than the central result. The other spatial results, such as angle between sources, surface normal, and source brightness ratios can also be seen as concurrent with colour rather than as predecessors.

Further, within brightness constancy, some spatial information can be inferred directly from the flux space computations. In particular it is easy to identify when all of the edge ratio is due to reflectance changes, barring coincidental alignment. Unfortunately, Gilchrist's False Mondrian demonstration depends on just such an alignment. Each region is carefully illuminated in such a way that every pixel has the same flux. This in itself is a degeneracy which perverts the flux space image to look as if there was only one source for the entire scene. Each flux structure is a single point, which also suggests that they are planes lit from a distance, and the lack of shadows suggests that they are coplanar.

In addition to the contrived alignment in the flux space, this illumination creates a degeneracy in the edges by making all of them of constant flux. This means that the ratio across the edge does not vary along its length. Any good edge detector would note the absence of the ratio changes inherent in occlusions, and this combined with the flux space results implies that the regions are coplanar, and so the flux ratio can be trusted as the value of the reflectance ratio. The eye is probably invoking the heuristic, based on the principle of minimum number of sources, that differences are attributed to

reflectance changes in the absence of contradictory data. Despite the superficial impressiveness of his Mondrian demonstration, it actually shows little more than does projecting a transparency on a white screen to demonstrate that perceived surface colour can be completely controlled by unnatural control of the lighting. The difference is that his "screen" was a patchwork of coloured pieces, separated in the third, unperceived, dimension for convenience in controlling their individual lighting.

Gilchrist's Mondrian would be more interesting if it was known when the illusion fails. As clues of different types are added to the scene, when does the eye realise that the regions are under different illumination? There are three projects here. Firstly, try to enhance the scene without changing the perception; at the moment the illusion has been shown only for regions of constant flux. Secondly, provide flux space clues without altering the edges, in order to test if human perception can make use of flux space information. Thirdly, provide edge clues without flux space clues, to see if edge information is sufficient. Scenes are now proposed for the latter two experiments.

The requirement of minimal edge information can be divided into two cases: edges with constant flux and edges with constant ratio. Constant flux is easier to create. The goal is regions which have non-constant overall flux, yet have a constant edge flux. This can be solved with an undulating Mondrian which has all the edges of each region coplanar, while the centre may bulge and dip. If these regions are substituted into Gilchrist's Mondrian, then each region will have a non-point flux structure yet a constant edge. It is likely that the undulations will make it easy to detect any

- 193 -

discrepancies in the angles of the sources, that is if one is lit from the top while another is lit from the side. Assuming this is the case, relight each area with parallel sources. If the undulations of each region are the same, it will probably allow direct flux comparisons. Therefore make the undulations of different patterns and measurements. Thus reduced, there should be nothing but flux space clues. The algorithms presented in the present work would only be partially successful because there is only one source per region and the edges provide incorrect brightness information. From this stage, try two alterations. One is to disrupt the incorrect edge clues to see if the flux structure is able to work when not given the wrong brightness ratio. The other is to keep the edges constant, but to add a second and third source to some or all the regions, where the new sources are also parallel but collectively from a different direction. The eye will probably notice when regions have a different number of sources. When each region has the same number of sources, flux space theory would predict that the situation where the two sources on one region do not have the same colour relationship as the two sources on another is detectable.

Using this, it should be possible to set up a conflict between the edge and flux space information. With one source colour, a region is made to seem a false colour by changing its illumination colour or brightness while making a complementary change in its reflectivity; the product is the same flux. Say the illumination has enriched blue. Now say that each region's illumination is divided into two sources, of differing colour, but which summed to the original light. If the false region's sources each have enriched blue by the original proportion, then the same transformation equates them, and the false

perception is unaltered or even strengthened. In contrast, say that the sum of the sources is unaltered, but that they are disproportionate to the sources of the neighbour, perhaps one has less blue while the other is greatly enhanced in blue. In this case, the flux space transform required to make the regions coplanar is different, and although the edge points are still of constant flux, they are shifted sideways in the flux plane and can not be equated for any brightness value. Thus there is an irreducible conflict between the edge evidence and the region evidence, which should produce an important result.

In this scene, if the spatial undulation is made big enough to produce a shadow in each colour in each region, then the width of the flux angle between the sources is determined as well as the plane they define. This produces a further conflict, because the diamonds are explicitly misaligned; in the previous case a colour transformation was possible but a brightness scalar was not, in this case even the colour transformation is not possible.

All of this has been done with a constant edge flux. To retry the important steps with a constant edge ratio requires even more careful setup. The goal is to produce a gradient which is shared along the edge, so that the flux is non-constant but the ratio across the edge is constant. Intuitively, this is stronger evidence that the regions are co-planar than simply a constant flux. It would be interesting to see if the eye agrees. The most obvious way to achieve the constant ratio is to use a distance gradient, again with each source parallel between regions, but now to have each also a constant distance from its region.

The complementary set of experiments is to provide edge

information without flux space information. As such, this is not possible, because whereas the edge may be held constant while the rest of the region varies, if the edge varies, then the flux of the region must vary at least as much because the edge is part of the region. One type of solution to this dilemma is to produce a scene which seems to have the same source or sources on each region, but does not have co-planar regions. It would appear to have flat regions which were slightly tilted so that their normals were different, and so that their distance gradients do not match. There will be difficulties in simultaneously matching the perpendicular edges, and also in elimination or correct treatment of shadows.

The implication in this design is that the colour system is not dependent on the entire spatial perception, as Gilchrist claims, but only on determining whether the regions share the same source colours and brightnesses. To this end, it would be interesting to set up his Mondrian so that it is clear that each region is independently suspended in space, yet the same false perception is obtained. The spatial clues could be shadows falling across different regions, or changing the normals of some planes to give different distance gradients. The illusion of a common source could be obtained by grouping the sources at one point, but each being a spotlight with a different filter. This would show that even with correct spatial perception the colours need not be correctly perceived; colour constancy uses the perceived spatial arrangement of the sources more than the spatial arrangement of the scene.

The other side of the coin is to show that colour constancy can be correct while the spatial perception of the scene is wrong. A trivial way to do this is to remove the filters from Gilchrist's own

scene. With all the regions now under the same light, the colours will all be seen correctly. However, the spatial arrangement will still be incorrectly interpreted as a coplanar Mondrian.

Land's experiments with a flash of colour could be extended. Their basis was that a previously unseen Mondrian was briefly illuminated by a flash of previously unseen colour. We have seen that it is easier to extract surface colours from Mondrian worlds. It would be interesting to see if colour constancy was as quick in more natural scenes. If a complex world with shadows and multiple illumination colours and angles could not always be solved correctly in a single flash, then there must be different levels of scene complexity. These could be isolated: occluding edges, shadows, multiple source angles, and multiple source colours.

It would also be interesting to make a scene of recognizable common objects which have known colours. However, modify some of the object colours slightly, and determine when the difference is great enough to be noticed. There are two questions here. We would expect colour constancy to have precedence over object recognition, so that a bluish banana would be recognised as such, rather than as a yellow banana, or as a bluish elongated object. This could be verified.

The other question is about tolerance bands under different illuminations. To a certain extent we perceive things as we know them to be. If the illumination is unfavourable, then we seem to permit a wider band of error before we think that the object has unexpected characteristics. Measuring the bands for different colours of objects under different illuminations may produce evidence about the human colour constancy mechanism, although instead it might be an artifact of the sensitivities of the cones and rods.

- 197 -

The latter possibility is based on a hypothesis about inherent perceptual bands which will now be stated. In the spectrum, the wavelengths vary continuously across its length; that is, each distance measures the same amount of wavelength change. However, humans perceive the spectrum differently, seeing a band of redish wavelengths followed by a band of orangeish; the difference between two reds is seen as smaller than an equal difference between a red and an orange. There is general agreement that these bands are fairly constant between individuals with normal colour vision. We would suggest that these bands are artifacts of the perceptual channels. A change of colour band may happen at the point where two receptors are giving equal output; to one side one receptor is greater, and to the other the other is greater. Several other functions are also candidates for colour change points. For example, it could be where a receptor is greater than the sum of the others, or where a receptor reaches its peak receptivity, and thus changes the slope of its receptivity curve.

Another interesting set of experiments would test the dimensionality of human colour vision. There is still some question as to whether there are exactly three types of cone, and there is certainly an issue of what roll, if any, the rods play. The flux space techniques give a method of determining the number of independent source colours in a scene, up to the number of receptors active in the colour algorithm. We would like an experiment which tests the ability to determine the number of different coloured sources in the scene. This is complicated by several factors.

The first is that the colours must be independent with respect to the human receptors. For two sources this is not hard, but for three

and four it would become increasingly difficult. A good initial solution would be Land's Retinex filters, which are each meant to simulate the receptivity of one of the four receptor channels.

The second complication is separating the number of sources from the number of source colours. That is, if sources illuminate the scene from left, right, top and bottom, then the gradients and spatial clues may well indicate that there are four sources. It is only after this that the subject might have a guess as to whether the sources were of the same or differing colours. A solution may be to place one source from very near the view of the observer, in order to produce the flattest lighting by it. This however, might produce anomalous results; recall that a uniform light behaves like background, and produces an offset from black for all the flux structures without increasing their dimensionality. While it would be very interesting to see if the eye distinguishes the background from a fully dimensional source, for example by noticing only the latter, this is not the question central to this experiment.

Associated with the problem of distinguishing sources before their colour is the problem that certain shadows may be thought to be giving too much evidence. For example, a free standing object on a plane will generate one shadow for every source, and these will probably overlap. If two sources are of the same colour, their shadows will also be. The number of sources, and which of them have the same colour can be immediately determined without substantial processing. It is fairly easy to reduce this problem by arranging the scene so that the shadows do not fall on the same colour of surface. In this way the reflectance change makes it impossible to compare their fluxes directly; although the shadows may still be used

in processing, this could happen only after colour constancy has compensated for the surface colours.

It is also easy to eliminate the shadows entirely, in order to determine whether they are necessary. This can be done with either a Mondrian or an undulating Mondrian.

An interesting and easier enhancement of the question of dimensionality in human vision distinguishes between local and global dimensionality. Imagine that five sources of independent colour light a scene, but there are never more than two sources on any region. The sources are set up so that they always overlap, like a Venn diagram, or like the links of a circular chain. Now in each part of the scene, only two dimensional colour processing would be necessary, but when these parts are linked, they fully permeate the humanly perceivable colour space. It is very likely that the human perceptual system correctly processes the scene, and suitable modifications can be made to the flux space algorithms to extend them into this unlikely domain. Notice that the concept of five independent sources is not entirely appropriate, since independence must be with reference to the human receptors. It is chosen in order to make the global perception beyond the abilities of the system.

It would be interesting to test the dimensionality of certain types of colour blindness.

Recall from the discussion of the role which rods play in colour vision, the observation that it is rare to perceive two different colours simultaneously at a single pixel. It is well known that if the perception is different between two stereo views, then it is usually seen as spectral reflection. However, the phenomenon we want here is monocular. It is not clear if it can be produced, or how to

- 200 -

approach the problem, but the attempt could provide additional insight into human colour perception mechanisms.

Envelope theory predicts that relative source brightness can be determined in a single region under three sources. Illuminate a region from three different angles by three independent colours. Have the curves of the surface show a range of flux in each colour. However, the brightest light should not give the brightest flux. This is achieved by having the surface turn towards the other sources, but never fully towards the brightest. The human observer should still be able to correctly decide which source was the brightest. If this is not the case, then it is another example where people can build a machine to do something which they can not do themselves.

The alternative scalings of the colour cube raise issues in colour perception, and experiments can be designed to see if the eye uses one particular method of scaling. Remove all of the regions of a standard Mondrian which have a high reflectivity in any channel. If the virtual white scaling is the only scaling used, then the cube will be perceived as smaller, which is the sensation of a darker light, and each region being seen as more reflective than before. Similarly, if only the regions which are highly reflective in one channel are removed, then the virtual white paradigm predicts that the light will be seen as lacking in that channel while the surface colours will seem enriched. These experiments probably require constant illumination across the scene to remove brightness clues.

Testing the body-fit algorithm would require removing many of the colours of the scene. To shift the colour away from red it is not

enough just to remove the reds. Each point which is to the red side of the diagonal must be avoided, which means that the red is greater than the average of the channels for that region, and includes dark as well as light regions. Further, this is not enough. Recall that the occupied portion of the colour cube is a double cone. If half of this is removed, through its axis, the best fit for a circle around its middle is unchanged, because the diameter of the cone is still intact. It is necessary to remove some of the width of this as well. The experiment would be improved if some surfaces of unusual reflectance were used, which were outside the normal cone. To distinguish between the virtual white and body-fit algorithms, these new surfaces would not be the most reflective in the scene, yet they could be used to create a displaced perception of the colour cube. Ideally, rather than simply removing part of the cones, a full sized cone would be constructed at a different flux angle. The trouble involved in attempting to fool the body-fit algorithm commends it for use in artificial vision.

Here is another experiment which uses related techniques, but has a different purpose. If each region in the Mondrian is enhanced in blue, is this perceived the same as if the illumination is similarly changed? This is the corollary of Gilchrist's Rooms. However, we have seen in chapter six that the rooms can be explained by secondary reflectance, both through the mutual reflectance and through depth of the shadows, which are due to the measurable background. These and other spatial clues would be absent, and it is likely that the perception would be the same.

In summarising flux space and envelope theory, it is useful to describe the conceptual steps which were involved in their

development. This will supplement the previous material which concentrated on their capabilities and procedures, and on what information was available for extraction from the scene.

Perhaps the first relevant discovery was that the dimensionality or basis number of the flux indicates the extent of the regions and the number of sources. An early attempt to capture this information was a mask which measured dimensionality, but there were several problems with this. First, any edge would increase the dimension. A shadow edge would raise the dimension by one in the lit area. A reflectance edge would often raise the dimension to span the whole range of fluxes. Although this seemed useful, when implemented it proved hard to extract the edges from the mask output, except in simple scenes. It was hard to distinguish close edges, and corners were usually saturated and indiscernible. Furthermore, constant flux in any source colour reduces the dimensionality, and thus creates false edges within a region. This was especially common on flattish regions, which showed up as a patchwork of shadows.

This led to the idea to actually compute the basis. This way it would be possible to know which source was added or removed by a shadow edge. The problems with constant flux could be eliminated, and region boundaries might be distinguished better. This generated much more computation and algorithmic complexity, and there was also a problem with selecting a suitable canonical form for expressing a measured basis so that the output of adjacent masks could be compared.

This eventually caused the masks to be discarded all together. The whole region could now be treated as one object, and flux space was created. The computational complexity was greatly reduced.

While examing the properties of flux structures, it was discovered that they align themselves when one is transformed onto another. Thus it was possible to determine the colour ratio of surfaces from the transformation between them, and also the direction of each source vector on any structure. Further, it is possible to identify when two regions were not lit by the same colour of sources, because their flux structures did not align.

It soon became clear that a flux structure is a Cartesian space with the source vectors as axes. This allows the effects of the individual sources to be separated at each pixel. The pattern of illumination brightness is related to the location in the structure, and so envelope theory was invented to use one convenient domain of illumination types. It may well be that other domains produce interesting and useful envelopes of other shapes.

When extending envelope theory to three-source scenes, it was realised that the envelope is always two dimensional because the space of surface normals is only two dimensional. This means that with three or more independent sources the envelope is hollow and every pixel falls on the envelope. This removes the ambiguities of planar points in the area under the two-source envelope, when it is only possible to state a minimum size to the envelope. With three colours the envelope's shape and size can be completely determined. This gives brightness information, previously lacking in flux space theory, and so completes colour constancy for these scenes.

Next it was observed that changes in brightness due to distance occur only slowly within a region, and often slowly across the whole scene. Therefore, it is possible to distinguish these from the change of surface normal which the envelope measures, and this

increases the range of scenes appropriate to this method.   This   way
of   separating distance   and   normal gradients was used again at the
beginning of region grouping.

One of the most important notions in working on colour   constancy
is that in the absence of other evidence, it can be assumed that
colour changes are due to the surface, and brightness changes are due
to the lighting. This is not strictly true for the flux of a region
under multiple sources, but it holds after the   extraction to flux
space.   The reason that colour changes are attributed to surfaces is
that there are many more objects than sources; to postulate that a
change is due to lighting usually means postulating an additional
source. The reason that brightness changes are usually due to the
illumination is that regions usually have a constant reflectivity,
but their normal changes, along with distance from the source and the
incidence of shadows.   Colour is related to the object's material,
while brightness is related to the spatial arrangement.

Another key notion was that regions should have only a single vote
in cube scaling and when estimating brightness. This removes the
chronic problem of large objects biasing the results.

It was observed that the ratio of the sources in a region is
independent of the reflectance colour of the region, so that distance
gradients between the sources are easy to detect. Since the sources
always have different locations, this allows the difference in their
gradients to be eliminated.

In worlds which are not lit by distant point sources, a
rectangular envelope may be used to transfer brightness values from
one source to another. Thus the more source colours and locations,
the more likely was a region to face towards one of them, and so have

its true brightness correctly estimated.

When there are two or more sources, if the points on either side of an edge are in the same light then they will match up for some brightness value after their colour transformation. Therefore, the transform gives a good test of edges, and any which align have a known brightness ratio.

Gilchrist's rooms suggested that there would be clues for separating illumination and reflectance in the room. There were two successful candidates: mutual illumination and shadows. It became clear that if the additional flux from secondary reflectance is compared with the primary flux which caused it, then the colour change would be entirely due to the colour of the secondary reflecting surface. This immediately gave the colours of the source and the surface.

Next it was discovered that the brightness gradient into a corner from either side converges to the same value. This gives the hope that the brightness of the source can also be determined without much spatial processing.

In the shadows of Gilchrist's rooms, all of the light is indirect, and so the shadow can be compared with the general level of flux from the room. The more uniform the overall flux, the more reliable the result. Reliability might also be measurable by the amount of gradient inside the shadows themselves. This led to the mathematical tricks for manipulating scenes which have ideal background illumination. Although the background would never be ideal in interesting scenes, this idea could be developed into a statistical comparison of primary flux, which is an estimate of the background level, with the thickness and offset of the diamond in the background

direction. This may yield a rough initial estimate of region brightness.

It should be realised that the essence of both solutions to Gilchrist's rooms lies in being able to look directly at the source of an illumination. In the case of mutual reflection, the relevant source is the opposite surface. With shadows, the source is the general flux within the room, which can likewise be easily measured.

Repeated use of the virtual white method of scaling the colour cube eventually led to its reappraisal, and the discovery of the body-fit algorithm, which uses more of the available information and so seems more robust.

An early worry that the flux space methods would require too much precision in the input data was relieved when it was discovered that it allows a very effective error correction facility. It is likely that this facility can also reduce the problems caused by coarse quantisation of brightness, which is a characteristic of currently available imaging systems.

It was realised that the ability to separate the effects of the different sources provides a way to photometric stereo on scenes in motion, and without turning the sources on and off.

One of the principles of artificial intelligence research is that the process of implementing a procedure always improves the understanding of that procedure and usually leads to its refinement. That has certainly been the case with flux theory. It was originally discovered when the use of conventional masks collapsed under the weight of their own epicycles. Many portions of the present algorithm have been implemented separately. Unifying these portions would be a complex task, but it would remove much of the confusion

from the process of grouping regions and scaling the colour cube.

The implementation should probably begin with artificial pictures. These have the advantage that brightness gradients of different types can be specified or eliminated as desired. Even impossible scenes can be created when appropriate for testing the program. At a later stage, artificial noise can be added, and the picture subjected to quantisation error and non-linear flux response. Only after all this should it be applied to real pictures.

In this work, a large range of techniques have been suggested for determining colour and brightness constancy. It seems likely that an algorithm would be most effective when using many or all of the techniques in parallel, with exchange of partial results. The colour constancy is easy, it is the brightness constancy which requires the work. Brightness measurements would seem to be a matter of step by step attributing the brightness changes to the different possible causes, and so building up a model of the illumination. Simultaneously there is the process of accumulating regions into groups large enough to form colour cubes. An actual implementation may well resemble Waltz filtering in the brightness constancy stage.

Bibliography

Abbreviations:

AIPC	International Conf on Automated Inspection and Product Control
ARVO	The Association for Research in Vision and Opthalmology
ASSP	Transactions on Acoustics, Speech, and Signal Processing
CGIP	Computer Graphics and Image Processing (through 1982)
CVGIP	Computer Vision, Graphics and Image Processing (from 1983)
IEEE	Institute of Electrical and Electronics Engineers
IJCAI	International Joint Conference on Artificial Intelligence
IOVS	Investigative Opthalmology & Visual Science
JOSA	Journal of the Optical Society of America
MIT	Massachusetts Institute of Technology
PAMI	Transactions on Pattern Analysis and Machine Intelligence
RoViSeC	International Conference on Robot Vision and Sensory Controls
SPIE	Society of Photo-Optical Instrumentation Engineers
SRI	Stanford Research Institute (International)

Akita, K. & Kuga H. (1979) "Towards Understanding Color Ocular
Fundus Images", **Proc of the Sixth IJCAI**, Tokyo, pp7-12.

Akita, K. & Kuga, H. (1982) "A Computer Method of Understanding
Ocular Fundus Images", **Pattern Recognition 15**:6 pp431-443.

Ali, M., Martin, W.N. & Aggraval, J.K. (1979) "Color-Based Computer
Analysis of Aerial Photographs", **CGIP 9**:3 pp282-293.

Aus, H.M., Harms, H., Huche, M., Gerlach, B. & Kriete, A. (1983)
"Computer Color-Vision", in **Proc of the Third RoViSeC**, eds. D.P.
Casasent & E.L. Hall, published as **Proc of the SPIE 449** pp225-229.

Barrow, H.G. & Tenenbaum, J.M. (1978) "Recovering Intrinsic Scene

Characteristics from Images", SRII Technical Note 157.

Also in **Computer Vision Systems**, eds. A.R. Hanson & E.M. Riseman (Academic Press, New York) pp3-26.

Bello, F. (1959) "An Astonishing New Theory of Color", **Fortune 59**:5 pp144-206.

Benton, J.L. & McCann, J.J. (1977) "Variegated Color Sensations from Rod-Cone Interactions: Flicker-Fusion Experiments", **JOSA 67**:1 pp119-121.

Bergstrom, S.S. (1977) "Common and Relative Components of Reflected Light as Information about the Illumination, Colour and Three-Dimensional Form of Objects", **Scandinavian Journal of Psychology 18**:3 pp180-186.

First published as Uppsala Psychological Report no.204.

Blackwell, H.R. & Blackwell, O.M. (1961) "Rod and Cone Receptor Mechanisms in Typical and Atypical Congenital Achromatopsia", **Vision Research 1**:1/2 pp62-107.

Brady, M. & Yuille, A. (1984) "An Extremum Principle for Shape from Contour", **IEEE Trans on PAMI 6**:3 pp288-301.

Ballard, D.H. & Brown, C.M. (1982) **Computer Vision**, (Prentice-Hall, Englewood Cliffs, New Jersey).

Brill, M.H. (1978) "A Device Performing Illuminant-Invariant Assessment of Chromatic Relations", **Journal of Theoretical Biology 71**:3 pp473-478.

Erratum **78**:2 page 308.

Brill, M.H. (1979) "Further Features of the Illuminant-Invariant Trichromatic Photosensor", **Journal of Theoretical Biology 78**:2 pp305-308.

Brill, M.H. & West, G. (1981) "Contributions to the Theory of

Invariance of Colors under the Conditions of Varying

Illumination", **Journal of Mathematical Biology** 11:3 pp337-.

Brill, M.H. (1984) "Physical and Information Constraints on the

Perception of Transparency and Translucency", **CVGIP** (in press).

Buchsbaum, G. & Goldstein, J.L. (1980) "Estimate of the Illuminant

in Spatially Complex Visual Fields: a Necessary Link in Accounting

for Colour Constancy", **Physics in Medicine and Biology 25:5**

page 1003. Meeting Abstract.

Chen, M.J. & Milgram, D.L. (1982) "Binary Color-Vision", in **Proc of**

the Second RoViSeC, Stuttgart (IFS Publications, Bedford)

pp293-306.

Cohen, P.R. & Feigenbaum, E.A. (1982) **The Handbook of Artificial**

Intelligence, Volume III, (William Kaufmann, Los Altos,

California).

Coleman Jr, E.N. & Jain, R. (1981a) "Obtaining 3-Dimensional Shape

of Textured and Specular Surfaces using Four-Source Photometry",

CGIP 18:4 1982 pp309-328. Originally Technical Report CSC-81-020

Wayne State Uni, Detroit.

Coleman Jr, E.N. & Jain, R. (1981b) "Shape from Shading for

Surfaces with Texture and Specularity", in **Proc of the Seventh**

IJCAI, Vancouver, pp652-657.

Connah, D.M. & Fishbourne, C.A. (1981) "The Use of Colour

Information in Industrial Scene Analysis", in **Proc of the First**

RoViSeC, Stratford-on-Avon (IFS Publications, Bedford) pp340-347.

Cornwell-Clyne, A. (1951) **Colour Cinematography**, (Chapman & Hall,

London). Expanded version of the 1936 original, which gives the

author's name as A.B. Klein.

Davis, L.S. & Rosenfeld, A (1981) "Cooperating Processes for Low-

level Vision: A Survey", **Artificial Intelligence 17** pp245-263.
Entire Volume reprinted as **Computer Vision,** ed. J.M. Brady
(North-Holland, Amsterdam).

Davis, L.S., Wang, C. & Xie, H. (1983) "An Experiment in
Multispectral, Multitemporal Crop Classification Using Relaxation
Techniques", **CVGIP 23:**2 1984 pp227-235.

Daw, N.W. (1962) "Why After Images are Not Seen in Normal
Conditions", **Nature 196:**4860 pp1143-1145.

Daw, N.W. (1973) "Neurophysiology of Color Vision", **Physiological
Reviews 53:**3 pp571-611.

Duda, R.O. & Hart, P.E. (1973) **Pattern Classification and Scene
Analysis,** (John Wiley, New York).

Eklundh, J.O., Yamamoto, H. & Rosenfeld, A. (1980) "A Relaxation
Method in Multispectral Pixel Classification", **IEEE Trans on PAMI
2:**1 pp72-75.

Falconer, D.G., Barrett, P. & Kottler, M. (1979) "Digital Analysis
of Color Retinal Photographs", SRII Technical Note 192.

Faugeras, O.D. (1979) "Digital Color Image Processing Within the
Framework of a Human Visual Model", **IEEE Trans on ASSP 27:**4
pp380-393.

Fiorini, R. & Montevecchi, F.M. (1982) "Automatic Colour Sensing",
in **Proc of the Sixth AIPC,** Birmingham, pp167-172.

Flock, H.R. (1984) "Illumination: Inferred or Observed", **Perception
and Psychophysics 35:**3 page 293.

Fox, W.F. & Hickey, W.H. (1914) "Improvements in Kinematographic
Apparatus", British Patent 636.

Frankle, J.A., Stiehl, W.A. & McCann, J.J. (1980) "Influence on
Lightness of Intraocular Scattered Light from White Surrounds",

ARVO 80, supplement to **IOVS,** page 212.   Meeting abstract.

Frisby, J.P. (1979) **Seeing: Illusion, Brain and Mind,** (Oxford
    University Press, Oxford).   Especially Chapter 6: "Seeing
    Lightness and Brightness", pp123-140.

Gilchrist, A.L.   (1975a) **Perceived Achromatic Color as a Function of**
    Ratios within Phenomenal Planes, Ph.D. Thesis, Rutgers University.

Gilchrist, A. & Rock, I. (1975b) "Achromatic Color Perception as a
    Function of Luminance Ratios within Phenomenal Planes", **Bulletin**
    of the Psychonomic Society 6:4B page 418.   Meeting abstract.

Gilchrist, A.L. (1977a) "Perceived Lightness Depends on Perceived
    Spatial Arrangement", **Science 195:**4274 pp185-187.

Gilchrist, A. & Delman, S. (1977b) "Discrimination of Illumination
    Edges from Reflectance Edges in Katz Paradigm", **Bulletin of the**
    Psychonomic Society 10:4 page 260.   Meeting abstract.

Gilchrist, A.L. (1977c) "Discrimination of Illumination Edges from
    Reflectance Edges", **ARVO 77,** supplement to **IOVS,** page 160.
    Meeting abstract.

Gilchrist, A., Jacobsen, A. & Delman, S. (1978) "Color Perception
    in a Single-Color Room", **ARVO 78,** supplement to **IOVS,** page 288.
    Meeting abstract.

Gilchrist, A.L. (1979a) "The Perception of Surface Blacks and
    Whites", **Scientific American 240:**3 pp112-124.

Gilchrist, A.L., Jacobsen, A. & Delman, S. (1979b) "Lightness
    Constancy with Changing Ratios", **ARVO 79,** supplement to **IOVS,**
    page 3.   Meeting abstract.

Gilchrist, A.L. (1980a) "When does Perceived Lightness Depend on
    Perceived Spatial Arrangement?", **Perception and Psychophysics 28:**6

pp527-538.

Gilchrist, A.L. (1980b) "Perception of Illumination and Surface
Reflectance based on Gradients of Light", **JOSA 70**:12 pp1562-1563.
Meeting abstract.

Gilchrist, A., Delman, S. & Jacobsen, A. (1983a) "The
Classification and Integration of Edges as Critical to the
Perception of Reflectance and Illumination", **Perception and
Psychophysics 33**:5 pp425-436.

Gilchrist, A.L. & Jacobsen, A. (1983b) "Lightness Constancy Through
a Veiling Luminance", **Journal of Experimental Psychology: Human
Perception and Performance 9**:6 pp936-944.

Gilchrist, A. & Jacobsen, A. (1984) "Perception of Lightness and
Illumination in a World of One Reflectance", **Perception 13**:1
pp5-19.

Hall, C.F. & Andrews, H.C. (1978) "Digital Color Image Compression
in a Perceptual Space", in **Applications of Digital Image
Processing**, San Diego, ed. A.G. Tescher, **Proc of the SPIE 149**
pp182-188.

Hamer, R.D., Alexander, K.R. & Teller, D.Y. (1982) "Rayleigh
Discriminations in Young Human Infants", **Vision Research 22**:5
pp575-587.

Hochberg, J.E. & Beck, J. (1954) "Apparent Spatial Arrangement and
Perceived Brightness", **Journal of Experimental Psychology 47**:4
pp263-266.

Holla, K. (1982) "Opponent Colors as a 2-Dimensional Feature within
a Model of the First Stages of the Human Visual System", in **Proc
of the Sixth Int Conf on Pattern Recognition**, Munich (IEEE

Computer Soc, Silver Spring MD) pp561-563.

Horn, B.K.P. (1973) "On Lightness", MIT AI Lab Memo no.295.

Horn, B.K.P. (1974) "Determining Lightness from an Image", **CGIP 3**:4
pp277-299.

Horn, B.K.P. (1975) "Obtaining Shape from Shading Information", in
The Psychology of Computer Vision, ed. P.H. Winston (McGraw-Hill,
New York) pp115-155.

Horn, B.K.P. (1978) "The Position of the Sun", MIT AI Lab Working
Paper 162.

Horn, B.K.P., Woodham, R.J. & Silver, W.M. (1978) "Determining
Shape and Reflectance using Multiple Images", MIT AI Lab Memo
no.490.

Horn, B.K.P. (1980) "Exact Reproduction of Colored Images", MIT AI
Lab Working Paper 207. Also in **CVGIP 26**:2 1984 pp135-167.

Hunt, B.R. & Kubler, O. (1984) "Karhunen-Loeve Multispectral Image
Restoration, Part I: Theory", **IEEE Trans on ASSP 32**:3 pp592-600.

Hurvich, L.M. & Jameson, D. (1960) "Color Vision", **Annual Review of
Psychology 11** pp99-130.

Hyde, P.D. & Davis, L.S. (1983) "Subpixel Edge Estimation", **Pattern
Recognition 16**:4 pp413-420.

Ito, T. (1973) "Towards Color Picture Processing", **CGIP 2**:3/4
pp347-354.

Ito, T. (1975) "Colour Picture Processing by Computer", in **Advance
Papers of the Fourth IJCAI**, Tbilisi, USSR, pp635-642.

Jameson, D. & Hurvich, L.M. (1961) "Complexities of Perceived
Brightness", **Science 133**:3447 pp174-179.

Jameson, D. & Hurvich, L.M. (1964) "Theory of Brightness and Colour

Contrast in Human Vision", **Vision Research** 4:1/2 pp135-154.

Jameson, D. & Hurvich, L.M. (1983) "Color Processing beyond
Mondrians", **JOSA 73**:12 page 1885. Meeting abstract.

Jarvis, R.A. (1982) "Expedient 3-D Robot Colour-Vision", in **Proc of
the Second RoViSeC,** Stuttgart (IFS Publications, Bedford)
pp327-338. Reprinted as "Expedient Range Enhanced 3-D Robot
Colour Vision", **Robotica 1**:1 1983 pp25-31.

Judd, D.B. (1960) "Appraisal of Land's Work on Two-Primary Color
Projections", **JOSA 50**:3 pp254-268.

Kanade, T. (1981) "Recovery of the Three-Dimensional Shape of an
Object from a Single View", **Artificial Intelligence 17** pp409-460.
Entire Volume reprinted as **Computer Vision,** ed. J.M. Brady
(North-Holland, Amsterdam).

Karp, A. (1959) "Colour-Image Synthesis with Two Unorthodox
Primaries", **Nature 184**:4687 pp710-712.

Karp, A. (1960) "Tritanopia and Two-Colour Image Synthesis", **Nature
188**:4744 pp40-42.

Kato, I, et. al. (1975) "A New Integrated Robot-Eye for Colour
Discrimination", **Fifth Symp on Industrial Robots,** (Soc of
Manufacturing Engineers, Dearborn) pp127-134.

Katz, D (1930) **Der Aufbau der Farbwelt.** Translated as **The World of
Colour** by R.B. MacLeod & C.W. Fox (Paul, London: 1935).

Kender, J. (1976) "Saturation, Hue and Normalized Color:
Calculation, Digitization Effects and Use", Tech Report, Dept Comp
Sci, CMU.

Kender, J.R. (1977) "Instabilities in Color Transformations", **IEEE
Conf on Pattern Recognition and Image Processing,** Rensselaer

University, Troy, New York, pp266-274.

Kittler, J. & Foglein, J. (1984) "Contextual Classification of
Multispectral Pixel Data", **Image and Vision Computing** 2:1 pp13-29.

Kolers, P.A. & von Grunau, M. (1975) "Visual Construction of Color
is Digital", **Science** 187:4178 pp757-759.

Kunii, T.L., Weyl, S. & Tenenbaum, J.M. (1974) "A Relational Data
Base Schema for Describing Complex Pictures with Color and
Texture", SRI Technical Note 93.

Also published in **Proc of the Second Int Joint Conf on Pattern
Recognition**, Lyngby-Copenhagen.

Land, E.H. (1958) "On the Nature of Color in the Image Situation",
JOSA 48:11 pp865-866. Meeting abstract.

Land, E.H. (1959a,b) "Color Vision and the Natural Image",
Proceedings of the National Academy of Science: Part I 45:1
pp115-129; Part II 45:4 pp636-644.

Land, E.H. (1959c) "Experiments in Color Vision", **Scientific
American** 200:5 pp84-99.

Land, E.H. (1959d) "Letter", **Scientific American** 201:3 pp16-22.

Land, E.H. (1960) "Some Comments on Dr. Judd's Paper", **JOSA** 50:3
page 268.

Land, E.H. (1962a) "Colour in the Natural Image", **Proceedings of
the Royal Institution of Great Britain** 39 pp1-15.

Land, E.H. & Daw, N.W. (1962b) "Colors Seen in a Flash of Light",
Proceedings of the U.S. National Academy of Science 48:6
pp1000-1008.

Land, E.H. (1964) "The Retinex", **American Scientist** 52:2 pp247-264.

Land, E.H. (1965) "The Retinex", in **Ciba Foundation Symposium on
Physiology and Experimental Psychology of Colour Vision**, eds.

G.E.W. Wolstenholme & J. Knight (J & A Churchill, London) pp217-223.

USA edition is **Colour Vision: Physiology and Experimental Psychology**, eds. A.V.S. DeReuck & J. Knight (Little & Brown, Boston).

Land, E.H. & McCann, J.J. (1971) "Lightness and Retinex Theory", **JOSA 61**:1 pp1-11, plus transparencies.

Land, E.H. (1974a) "The Retinex Theory of Colour Vision", **Proceedings of the Royal Institution of Great Britain 47** pp23-58.

Land, E.H. (1974b) "Smitty Stevens' Test of Retinex Theory", in **Sensation and Measurement. Papers in Honor of S.S. Stevens**, eds. H.R. Maskowitz, B. Scharf & J.C. Stevens (D. Reidel, Dordrecht Holland) pp363-368.

Land, E.H. (1977a) "6 Eyes of Man", in **Seminar on 3-Dimensional Imaging**, San Diego, ed. S.A. Benton, **Proc of the SPIE 120** pp43-51.

Land, E.H. (1977b) "The Retinex Theory of Color Vision", **Scientific American 237**:6 pp108-128.

Land, E.H. (1979) "Role of the Retinex", in **Los Alamos Conf on Optics**, ed. D.H. Liebenberg, **Proc of the SPIE 190** page 484.

Land, E.H. (1981) "Vision and Color", in **Encyclopedia of Physics**, eds. R.G. Lerner & G.L. Trigg (Addison-Wesley, Reading Massachusetts) pp1090-1097.

Land, E.H., Hubel, D.H., Livingstone, M.S., Perry, S.H. & Burns, M.M. (1983a) "Colour-Generating Interactions across the Corpus Callosum", **Nature 303**:5918 pp616-618.

Land, E.H. (1983b) "Recent Advances in Retinex Theory and some
  Implications for Cortical Computations - Color Vision and the
  Natural Image", **Proceedings of the National Academy of Sciences of
  the USA - Physical Sciences** 80:16 pp5163-5169.

Lettvin, J.Y. (1967) "The Colors of Colored Things", Quarterly
  Progress Report 87, MIT Research Lab of Electronics, pp193-229.

Linden, L.L. & Lettvin, J.Y. (1975) "Evolutionary Constraints on
  the Dimensionality of Color Space", MIT Research Lab of
  Electronics, Progress Report 116, pp288-302.

Linden, L. & Lettvin, J. (1977) "Freedom and Constraints in Color
  Vision", **Brain Theory Newsletter** 3:2 pp29-30.

Loughlin, C. (1982) "INSPECTRUM - Full Colour, High Resolution
  Inspection System", in **Proc of the Sixth AIPC**, Birmingham,
  pp135-144.

MacLeod, R.B. (1932) "An Experimental Investigation of Brightness
  Constancy", **Archives of Psychology 135.**

MacVicar-Whelan, P.J. & Binford, T.O. (1981) "Intensity
  Discontinuity Location to Subpixel Precision", in **Proc of the
  Seventh IJCAI**, Vancouver, pp752-755.

Marks, W.B., Dobelle, W.H. & MacNichol Jr., E.F. (1964) "Visual
  Pigments of Single Primate Cones", **Science** 143:3611 pp1181-1183.

Marr, D. (1974) "The Computation of Lightness by the Primate
  Retina", **Vision Research** 14:12 pp1377-1388.

Marr, D. (1982) **Vision: A Computational Investigation into the
  Human Representation and Processing of Visual Information**, (W.H.
  Freeman, San Francisco).

McCann, J.J. & Benton, J.L. (1969) "Interaction of the Long-Wave
  Cones and the Rods to Produce Color Sensations", **JOSA** 59:1

pp103-107.

McCann, J.J., Land, E.H. & Tatnall, S.M.V. (1970) "A Technique for
Comparing Human Visual Responses with a Mathematical Model for
Lightness", **American Journal of Optometry and Archives of American
Academy of Optometry** 47:11 pp845-855.

McCann, J.J. (1972) "Rod-Cone Interactions: Different Color
Sensations from Identical Stimuli", **Science** 176:4040 pp1255-1257.

McCann, J.J. (1973) "Human Color Perception", in **Color: Theory and
Imaging Systems** (Soc of Photographic Scientists and Engineers,
Washington, D.C.) pp1-23.

McCann, J.J., Savoy, R.L. & Hall Jr.,J.A. (1973) "Visibility of
Low-Spatial-Frequency Sine-wave Targets: Dependence on Number of
Cycles", **JOSA** 63:10 page 1297. Meeting abstract.

McCann, J.J., Savoy, R.L., Hall Jr., J.A. & Scarpetti, J.J. (1974)
"Visibility of Continuous Luminance Gradients", **Vision Research**
14:10 pp917-927.

McCann, J.J. & Benton, J.L. (1975) "Variegated Color Sensations
from Rod-Cone Interactions: Flicker-Fusion Experiments", **JOSA**
65:10 page 1200. Meeting abstract.

McCann, J.J., McKee, S.P. & Taylor, T.H. (1976a) "Quantitative
Studies in Retinex Theory: A Comparison between Theoretical
Predictions and Observer Responses to the 'Color Mondrian'
Experiments", **Vision Research** 16:5 pp445-458.

McCann J.J. (1976b) "The Maximum Luminance in the Field of View
Does Not Always Generate the Maximum Lightness", **JOSA** 66:10
page 1103. Meeting abstract.

McCann, J.J., Hall Jr., J.A. & Land, E.H. (1977) "Color Mondrian
Experiments: The Study of Average Spectral Distributions", **JOSA**

67:10 page 1380. Meeting abstract.

McCann, J.J., Stiehl, W.A. & Savoy, R.L. (1979) "Influence of Intraocular Scattered Light on Lightness Scaling Experiments", **JOSA 69**:10 p1452. Meeting abstract.

McCann, J.J., Land, E.H., Frankle, J.A. & Stiehl, W.A. (1980) "Retinex Processing of Natural Images", **ARVO 80,** supplement to **IOVS,** page 135. Meeting abstract.

McCann, J.J. & Frankle, J.A. (1982) "Method and Apparatus for Lightness Imaging", European patent application 0046988.

McCann, J.J. & Houston, K.L. (1983a) "Color Sensation, Color-Perception and Mathematical Models of Color-Vision", in **Colour Vision: Physiology & Psychophysics,** eds. J.D. Mollon & L.T. Sharpe (Academic Press, London) pp535-544.

McCann, J.J. & Houston K.L. (1983b) "Calculating Color Sensations from Arrays of Physical Stimuli", **IEEE Trans on Systems, Man & Cybernetics 13**:5 pp1000-1007.

McKee, S.P., McCann, J.J. & Benton, J.L. (1977) "Color Vision from Rod and Long-wave Cone Interactions: Conditions in which Rods Contribute to Multi-Colored Images", **Vision Research 17**:2 pp175-185.

McLeod, R.D. (1978) "New Color Vision Model, including Color Blindness", **Biophysical Journal 21**:3 page 118a. Meeting abstract.

McLeod, R.D. (1978) "Colored Pincushion Grids, Complements, and Color Vision", **JOSA 68**:10 page 1400. Meeting abstract.

Mollon, J.D. (1982) "Color Vision", **Annual Review of Psychology 33** pp41-85.

Nagy, A.L., MacLeod, D.I.A., Heyneman, N.E. & Eisner, A. (1981) "Four Cone Pigments in Women Heterozygous for Color Deficiency",

JOSA 71:6 pp719-722.

Nevatia, R. (1977) "A Color Edge Detector and Its Use in Scene
Segmentation", **IEEE Trans on Systems, Man & Cybernetics** 7:11
pp820-826.

Nguyen, N.G., Poulsen, R.S. & Louis, C. (1983) "Some New Color
Features and Their Application to Cervical Cell Classification",
Pattern Recognition 16:4 pp401-411.

Ohlander, R.B. (1975) **Analysis of Natural Scenes**, Ph.D. Thesis,
Carnegie-Mellon University, Pittsburg.

Ohlander, R., Price, K. & Reddy, D.R. (1978) "Picture Segmentation
using a Recursive Region Splitting Method", **CGIP** 8:3 pp313-333.

Ohta, Y., Kanade, T. & Sakai, T. (1980) "Color Information for
Region Segmentation", **CGIP** 13:3 pp222-241.

Pearson D.E., Rubinstein, C.B. & Spivack, G.J. (1969) "Comparison
of Perceived Color in Two-Primary Computer-Generated Artificial
Images with Predictions based on the Helson-Judd Formulation",
JOSA 59:5 pp644-658.

Pentland, A.P. (1982a) "Finding the Illuminant Direction", **JOSA**
72:4 pp448-455.

Pentland, A.P. (1982b) "Local Shading Analysis", **IEEE Trans on PAMI**
6:2 1984 pp170-187. Originally SRI Tech Note 272.

Pentland, A.P. (1982c) "Perception of Shape from Shading", **JOSA**
72:12 page 1756. Meeting abstract.

Price, K.E. (1977) **Change Detection and Analysis in Multi-Spectral
Images**, Ph.D. Thesis, Dept Comp Sci, CMU.

Pulos, E., Teller, D.Y. & Buck, S.L. (1980) "Infant Color Vision: A
Search for Short-Wavelength-Sensitive Mechanisms by Chromatic

Adaptation", **Vision Research** 20:6 pp485-493.

Rebollo, M. & Escudero, L.F. (1977) "A Mixed Integer Programming Approach to Multi-Spectral Image Classification", **Pattern Recognition** 9:1 pp47-57.

Redding, G.M. & Lester, C.F. (1980) "Achromatic Color Matching as a Function of Apparent Target Orientation, Target and Background Luminance and Lightness or Brightness Instructions", **Perception and Psychophysics** 27:6 pp557-563.

Richards, W. (1975) "Mosaic Model for Color Vision", **Journal of Theoretical Biology** 53:1 pp177-197.

Robinson, G.S. (1977) "Color Edge Detection", **Optical Engineering.**

Rock, I. (1977) "In Defence of Unconscious Inference", in **Stability and Constancy in Visual Perception: Mechanisms and Processes,** ed. W. Epstein (Wiley, New York) pp321-373.

Savoy, R.L. & McCann, J.J. (1975) "Visibility of Low-Spatial-Frequency Sine-wave Targets: Dependence on Number of Cycles", **JOSA** 65:3 pp343-350.

Schachter, B.J., Davis, L.S. & Rosenfeld, A. (1976) "Scene Segmentation by Cluster Detection in Color Space", **ACM Sigart Newsletter** no.58 pp16-17.

Schaerf, R. & Mauer, E. (1982) "Evaluation of Multispectral Image by Feature Combination", in **Proc of the Sixth Int Conf on Pattern Recognition,** Munich (IEEE Computer Soc, Silver Spring MD) pp554-556.

Shirai, Y. & Tsuji, S. (1971) "Extraction of the Line Drawings of 3-D Objects by Sequential Illumination from Several Directions", **Pattern Recognition** 4:4 1972 pp343-351.

First printed in: **Second IJCAI,** 1971 pp71-79.

Smith, G.B. (1983) "Shape from Shading: An Assessment", SRII AI
    Center Technical Note 287.

Stabell, U. & Stabell, B. (1965) "Rods as Color Receptors",
    Scandinavian Journal of Psychology 6:3 pp195-200.

Stephens, B.R. & Banks, M.S. (1984) "Development of Contrast
    Constancy in Human Infants", **ARVO 84,** supplement to **IOVS,**
    page 162. Meeting abstract.

Stiehl, W.A., McCann, J.J. & Savoy, R.L. (1983) "Influence of
    Intraocular Scattered Light on Lightness-Scaling Experiments",
    JOSA 73:9 pp1143-1148.

Stromeyer III, C.F. (1971) "McCollough Effect Analogs of Two-Color
    Projections", **Vision Research 11:**9 pp969-978.

Tabatabai, A.J. & Mitchell, O.R. (1984) "Edge Location to Subpixel
    Values in Digital Imagery", **IEEE Trans on PAMI** 6:2 pp188-201.

Tenenbaum, J.M. (1973) "On Locating Objects by their Distinguishing
    Features in Multisensory Images", SRI Technical Note 84. Also
    published as **CGIP** 2:3/4 pp308-320.

Tenenbaum, J.M., Garvey, T.D., Weyl, S. & Wolf, H.C. (1974) "An
    Interactive Facility for Scene Analysis Research", SRI Final
    Report, Project 1187, Tech note 87.

Tenenbaum, J.M. & Weyl, S. (1975) "A Region-Analysis Subsystem for
    Interactive Scene Analysis", in **Advance Papers of the Fourth**
    IJCAI, Tbilisi, USSR,

Todd, J.A. (1947) **Projective and Analytical Geometry,** (Pitman &
    Sons, London).

Tou, J.T. & Gonzalez, R.C. (1974) **Pattern Recognition Principles,**

(Addison-Wesley, Reading, Massachusetts).

Ueda, M., Matsuda, F. & Sako, S. (1980) "Color Sensing System for
    an Industrial Robot", in **Proceedings - Tenth Int Symp on
    Industrial Robots**, Milan (IFS Publications, Bedford, England)
    pp153-162.

Ullman, S. (1976) "On Visual Detection of Light Sources",
    Biological Cybernetics 21:4 1976 pp205-212. Substantial revision
    of MIT AI Lab Memo no.333, 1975. Abbreviated in **Artificial
    Intelligence: An MIT Perspective, Volume 2: Understanding Vision,
    Manipulation, Computer Design, Symbol Manipulation**, eds P.H.
    Winston & R.H. Brown (MIT Press, Cambridge, Mass.: 1979) pp83-100.

Walker, J. (1979) "Experiments with Edwin Land's Method of Getting
    Color out of Black and White", **Scientific American 240**:6
    pp142-146.

Wallach, H. (1948) "Brightness Constancy and the Nature of
    Achromatic Colors", **Journal of Experimental Psychology 38**:3
    pp310-324.

Wallach, H. (1963) "The Perception of Neutral Colors", **Scientific
    American 208**:1 pp107-114.

Walls, G.L. (1960) "Land! Land!", **Psychological Bulletin 57**:1
    pp29-48.

Waltz, D. (1972) **Generating Semantic Descriptions from Drawings of
    Scenes with Shadows**, MIT AI Lab Technical Report 271.
    Reprinted as "Understanding Line Drawings of Scenes with Shadows",
    in **The Psychology of Computer Vision**, ed. P.H. Winston (McGraw-
    Hill, New York: 1975) pp19-91.

West, G. & Brill, M.H. (1982) "Necessary and Sufficient Conditions
    for Von Kries Chromatic Adaptation to Give Color Constancy",

Journal of Mathematical Biology 15:2 pp249-258.

Wheeler, S.G. & Misra, P.N. (1980) "Crop Classification with
Landsat Multispectral Scanner Data II", **Pattern Recognition 12:4**
pp219-228.

Wilmer, E.N. (1949) "Low Threshold Rods and the Perception of
Blue", **Journal of Physiology, London 111** page 17P.

Wilson, M.H. & Brocklebank, R.W. (1960) "Two-Colour Projection
Phenomena", **Journal of Photographic Science 8:** pp141-.

Wilson, M.H. & Brocklebank, R.W. (1961) "Colour and Perception: the
work of Edwin Land in the Light of Current Concepts", **Contemporary
Physics 3:2** pp91-111.

Woodham, R.J. (1978) "Photometric Stereo: A Reflectance Map
Technique for determining Surface Orientation from Image
Intensity", in **Seminar on Image Understanding Systems and
Industrial Applications**, San Diego, ed. R. Nevatia, **Proc of the
SPIE 155.**

Also available as MIT AI Lab Memo no.479.

Worthey, J.A. (1982) "Opponent-Colors Approach to Color Rendering",
JOSA 72:1 pp74-82.

Worthey, J.A. (1983) "Illuminant Centered View of Color Constancy",
JOSA 73:12 page 1902. Meeting abstract.

Wright, W.D. (1959) "Colour Vision: a Field of Unsolved Problems",
New Scientist 6:1 48 pp447-449.

Yachida, M. & Tsuji, S. (1971) "Application of Color Information to
Visual Perception", **Pattern Recognition 3:3** pp307-323.

Yoshimoto, K. & Torige, A. (1983) "Development of a Colour

Information Processing System for Robot Vision", in **Developments in Robotics 1983**, ed. Brian Rooks (IFS, Bedford) pp161-166.

Zeki, S. (1980) "The Representation of Colours in the Cerebral Cortex", **Nature** 284:5755 pp412-418, plus three retinex filters.

Zeki, S. (1983a) "Does the Colour of the After Image depend upon Wavelength Composition?", **Journal of Physiology, London** 340 page 49P.

Zeki, S. (1983b) "A Demonstration of 'Void' Colours and 'Natural' Colours", **Journal of Physiology, London** 341 page 1P.

Zeki, S. (1983c) "Colour Coding in the Cerebral Cortex: the Reaction of Cells in Monkey Visual Cortex to Wavelengths and Colours", **Neuroscience** 9:4 pp741-765.

Zeki, S. (1983d) "Colour Coding in the Cerebral Cortex: the Responses of Wavelength Selective and Colour-coded cells in Monkey Visual Cortex to Changes in Wavelength Composition", **Neuroscience** 9:4 pp767-781.