COLOUR OBJECT RECOGNITION

by

Graham David Finlayson

B.Sc University of Strathclyde 1989

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
in the School
of
Computing Science

SIMON FRASER UNIVERSITY
April 1992

APPROVAL

Name: Graham David Finlayson

Degree: Master of Science

Title of thesis: Colour Object Recognition

Examining Committee:

Dr. Jia-wei Han
Chair

Dr Brian V. Funt
Senior Supervisor

Dr Ze-Nian Li
Examiner

Dr Michael Swain
Department of Computing Science
University of Chicago
External Examiner

Date Approved: _____

Abstract

Since colour characterizes local surface properties and is largely viewpoint insensitive it is a useful cue for object recognition. Indeed, Swain and Ballard have developed a simple scheme, called colour-indexing, which identifies objects by matching colour-space histograms. Their approach is remarkably robust in that variations such as a shift in viewing position, a change in the scene background or even object deformation degrade recognition only slightly. Colour-indexing fails, however, if the intensity or spectral characteristics of the incident illuminant varies. This thesis examines two different strategies for rectifying this failure.

Firstly we consider applying a colour constancy transform to each image prior to colour-indexing (colours are mapped to their appearance under canonical lighting conditions). To solve for the colour constancy transform assumptions must be made about the world. These assumptions dictate the types of objects which can be recognized by colour-indexing + colour constancy preprocessing. We review several colour constancy algorithms and in almost all cases conclude that their assumptions are too limiting. The exception, a discrete implementation of Forsyth's CRULE, successfully solves the colour constancy problem for sets of simple objects viewed under constant illumination.

To circumvent the need for colour constancy preprocessing and to recognize more complex object sets we consider indexing on illuminant invariants. Three illuminant invariants—volumetric, opponent and ratio—are examined. Each characterizes local surface properties, is largely viewpoint insensitive **and** is independent of both the intensity and spectral characteristics of the incident illuminant. We develop an algorithm, called Colour constant colour-indexing, which identifies objects by matching colour ratio-space histograms. In general our algorithm performs comparably

with colour-indexing under fixed illumination, but substantially better than colour-indexing under varying illumination.

Acknowledgements

My sincerest thanks go to my supervisor Dr. Brian V. Funt. He has been exceptionally supportive throughout the duration of this thesis. In particular I greatly appreciate Dr. Funt's approach to research; the free exchange of ideas forms the foundation of a stimulating working environment.

Regarding research, I am also indebted to Dr. Mark Drew. His research skills and tireless enthusiasm are an example to all.

Special thanks go to my fellow graduate students in the Computing Science department. They contribute greatly to the general research environment and more importantly supply support and friendship. In particular I'd like to mention computing science stalwarts—Allan Bennet-Brown, Gilles Dionne and Mark Mezofenyi. In a category by himself is Sanjeev Mahajan—as well as being a close friend, he provided valuable insights concerning the probability results of this thesis.

My final thanks goes to the departmental advisor—Elma Krbavac. I met Elma whilst studying in Glasgow in May 1989. She was my introduction to the computing science department at Simon Fraser University. This meeting led to an adventure which is still continuing.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Colour plays an important role in object identification. For example an apple can be distinguished from an orange solely on the basis of colour. In this thesis we develop a machine vision system which can robustly identify colourful objects. Identification takes place in real time and as such would be be a useful part of an active vision system.

Machine vision systems can recognize objects only when certain assumptions are satisfied. The weakest assumption, common to all *model-based* vision approaches, is that objects can only be identified if they have been seen before. Further model-based vision systems are usually told, by a system designer, to remember particular objects. This implies there are two stages to identification. In the *learning* stage the vision system views the objects which it must identify. The images of objects are analysed, features extracted, and these features are grouped to form *canonical models*; the set of all canonical models is called the model database. In the *active* stage an object is presented to the vision system; again its image is analysed and an *image model* is built. The image model is then matched against the model database. The best canonical match identifies the image model.

Ideally the success of object identification should not require strong constraints on the world. For example, if objects are always presented to the vision system in the same orientation and at the same relative position then this contextual information can be exploited in designing a model representation. However strong constraints will

limit the utility of the vision system. It is desirable that identification performance be unaffected by the visual context in which an object is viewed. For the purposes of this thesis changes in visual context refer to:

1. changes in the background of an object.
2. occlusion of the object.
3. changing relative position of the object.
4. the lighting conditions:
   (a) changing light intensity.
   (b) varying spectral characteristics of the light.

Traditional approaches to object identification are based on single-channel intensity images and rely on geometric descriptions of objects. For example, the 3-dimensional shape of an object would obviously serve as a good key for identification. However, for unconstrained scenes, it is difficult to to extract this 3-dimensional information. Lower level geometric cues are often used: these include looking at edge intersections or at relative edge orientations[31]. Unfortunately there are few geometric cues which are invariant to changes in viewing context.

Swain[28] departs from the geometric approach and instead develops a simple scheme which identifies objects entirely on the basis of colour. His method, called **Colour-indexing**, is extremely successful at identifying objects and is largely unaffected by the first three changes in viewing context.

1.1 Swain's Colour-indexing

Swain's colour-indexing algorithm identifies an object by comparing its colours to the colours of each object in the model database (a colour refers to a response 3-vector *registered* by three sensor channels). The area of a particular colour is calculated and is stored as the bin-count of a 3-Dimensional histogram (the model), appropriately

called the *colour histogram.* Similar 3-vectors are mapped to the same histogram bins. Specifically each colour channel is discretised into 16 intervals; hence, each colour histogram has $16 * 16 * 16 = 4096$ bins.

Objects are presegmented from the images when calculating canonical models. This prevents background colours affecting object identification. In the active stage objects are not segmented from their backgrounds (as such segmentation necessarily implies knowledge of the object in the image!).

Histograms are matched by comparing the counts, or areas, in corresponding bins via a technique called *Histogram intersection.* The intersection of histograms H_1 and H_2 is defined as:

$$H_1 \bigcap H_2 \; = \; \sum_i \sum_j \sum_k \min(H_1(i,j,k), H_2(i,j,k)) \tag{1.1}$$

Since canonical histograms contain no background colours, intersection (or match) values are normalized by the number of pixels in the model histogram, thus matches are between 0 and 1. Histogram intersection is very fast requiring time proportional to the number of histogram bins. More sophisticated correlation measures could be used, but the success of colour indexing implies that they may not be necessary.

Let us examine the performance of colour-indexing with respect to the 4 changes in visual context:

1. Changing the background in which an object is imaged will only add to the match value of the histogram intersection if:

   (a) the pixel has the same colour as one of the colours in the model.

   (b) the number of pixels of that colour in the model is less than the number of pixels of that colour in the image.

   Thus the correct match will always be found unless the two objects are very similar or the background is specifically designed to confound matching.

2. Experimentally Swain demonstrates that histogram intersection continues to work well even when an object is partially occluded. This is to be expected as

a colour histogram is an accumulation of global evidence. The coloured areas not occluded should still be sufficiently discriminatory to allow correct object identification.

3. As the viewer alters position (or the object is rotated) some colours may come into view and others disappear. In this case the colour histograms can change quite significantly. Swain's solution to this problem is to store histograms for each model as seen from differing viewpoints.

   Another problem can occur as the viewing position changes. In a world of lambertian surfaces with point source illumination, the brightness of a surface changes as the angle between the lluminant vector and the surface normal varies. If $\underline{v}$ denotes the illuminant direction and $\underline{n}$ represents the surface normal then the brightness of the reflected light is proportional to $\underline{v}.\underline{n}$ (the vector dot-product). This relationship accounts for the shading field in an image. By implication the brightness, or magnitude, of colours in an image will change as an object is rotated relative to the illuminant. Again storing multiple colour histograms for each model viewed under varying conditions may help here.

   Normalizing the lengths of sensor response vectors, by dividing by a linear combination of the responses (the vector components), gives intensity-independent colours. Such a normalization results in 2-dimensional information (the normalized blue response can be generated given the normalized red and green responses). Swain[29] histograms response vectors normalized with respect to the sum of the red, green and blue responses. Color-indexing continues to perform relatively well despite this shift from a 3-dimensional to a 2-dimensional index. However a significant number of test images ($\approx 25\%$) are poorly matched. In section 3 we readdress this problem when designing an illuminant invariant feature space for object identification.

4. Altering the intensity of the light effectively alters the length of the colour tuples. That is every pixel in the image will be multiplied by a constant factor k. Swain presents experimental results for such intensity changes and concludes that even for values of k fairly close to 1 object identification is impaired. Changing the

spectral characteristics of the illuminant hamper Swain's algorithm to a greater degree—both the length and direction of colour vectors will change.

1.2 Extending Colour-indexing

Swain's Colour-indexing fails since the colours registered by a vision system are dependent on the relative position of the object (the shading field) and, more importantly, on the spectral characteristics of the illuminant. Swain proposes applying a colour constancy algorithm[10, 21, 12, 11] to the registered colours thereby removing the effects of a varying illuminant. Each registered colour is mapped to its appearance under canonical lighting conditions.

Unfortunately, even for fairly simple worlds, the general colour constancy problem is as yet unsolved. However the model-based identification problem imposes constraints on the world. In particular, since the model database contains a finite number of models, this implies that the world contains a finite set of surfaces. This constraint can be exploited in a discrete version of Forsyth's[10] CRULE algorithm. For simple worlds, where objects are 2-dimensional and where illumination is everywhere uniform, CRULE successfully solves the colour constancy problem and facilitates object identification.

In more complex worlds, where objects are 3-dimensional and illumination is allowed to vary, CRULE cannot solve the colour constancy problem. For this reason we develop a new approach to object identification called **Colour constant colour indexing**[1]. This scheme indexes not on colour triples but on illuminant invariants. In particular the ratio of two neighbouring colours is, more or less, illumination independent; colour ratios form the backbone of colour constant colour indexing.

[1] Colour constant colour indexing was jointly developed in collaboration with Dr. Brian V. Funt.

1.3 Thesis Overview

In chapter 2 we examine, in detail, the colour constancy problem. Many computational theories have been proposed; each of which places restrictions on the type of object for which Swain's colour indexing can work. In almost all cases these restrictions are not satisfied by any plausible object set. However we conclude chapter 2 by presenting a discrete version of Forsyth's CRULE colour constancy algorithm; CRULE can successfully solve the colour constancy problem for simple sets of objects.

Unfortunately for most realistic object sets the colour constancy problem cannot (as yet) be solved. To circumvent the need for colour constancy preprocessing we consider, in Chapter 3, identification based on illuminant invariants. Three types of invariants are considered: volumetric, opponent and ratio. Colour ratios have favourable error and computational properties. Consequently colour ratios form the backbone of a new identification algorithm— colour constant colour indexing.

Various representational issues result from the switch from colours to colour ratios. These are addressed in chapter 4. In particular we show that the distribution of colour ratios is non-uniform; this implies the bins of the *ratio histograms* should be of different size. Issues related to ratio error are also explored.

Chapter 5 presents experimental data contrasting the performance of colour indexing and colour constant colour indexing under illumination change. For sets of synthetic and real images, colour constant colour indexing is extremely successful at identifying objects; this contrasts with the poor identification success of colour indexing.

Chapter 2

Colour Constancy

The perceived colour of an object is, more or less, independent of the illuminant under which it is observed. Thus colour is a quality not of the reflected light but of the object's surface. The ability to label objects with colour names, that refer only to surface reflectance properties, is called colour constancy[2].

Humans have 3 types of colour receptors: long-, medium-, and short-wave sensitive cones. Hence the eye measures, at most, 3 properties of surface reflectances. Therefore colour constancy requires that the initial 3-vector of cone responses be transformed into a 3-parameter surface descriptor. The colour constancy problem in machine vision is similar; though, there is no restriction on the number of receptors.

Various algorithms have been proposed for solving the colour constancy problem. Each algorithm places restrictions on the types of surfaces and illuminants in the world. If the set of objects we wish to identify satisfies these restrictions then Colour Indexing+colour constancy preprocessing will successfully identify objects under varying lighting conditions.

In this section we review several existing colour constancy algorithms. In all cases their world restrictions are very strong; indeed these restrictions cannot be satisfied by any plausible object set. However consideration of the object identification problem yields new constraints and these are elegantly incorporated into a discrete version of Forsyth's CRULE algorithm. CRULE effectively solves the colour constancy problem for a set of 2-dimensional objects, where the total number of distinct colours is small,

viewed under constant illumination.

2.1 Preliminaries

The light reflected from a surface depends not only on the spectral properties of illumination and surface reflectance, but also on other confounding factors; these include specularities and mutual illumination. For this reason computational theories for colour constancy are often developed for the simplified Mondriaan world; a Mondriaan is a planar surface composed of several, overlapping, matte (lambertian) patches. The light striking the Mondriaan is assumed locally constant, i.e the intensity and spectral characteristics of the light varies slowly. In this world the only confounding process to retrieving surface descriptors is illumination. Almost all colour constancy algorithms are designed for the Mondriaan world.

A priori to examining any colour constancy algorithm, the Mondriaan assumption has severely restricted the types of objects which can be recognized. In particular, objects are constrained to be planar. There are no colour constancy algorithms which work in an unconstrained 3-dimensional world.

2.1.1 Sensor Responses

Light reflected from a Mondriaan falls onto a planar array of sensors, analogous to the retina. At each location x in the sensor array there are s different classes of sensors. The value registered by the kth sensor (a scalar), p_k^x, is equal to the integral of its response function multiplied by the incoming colour signal. Each p_k^x corresponds to a unique surface reflectance.

$$p_k^x = \int_\omega C^x(\lambda)\, R_k(\lambda)\, d\lambda \tag{2.1}$$

where λ is wavelength, $R_k(\lambda)$ is the response function of the kth sensor, $C^x(\lambda)$ is the colour signal at x and the integral is taken over the visible spectrum ω. The colour signal is the product of a single surface reflectance $S(\lambda)$ multiplied by the ambient illumination $E(\lambda)$, $C(\lambda) = E(\lambda)S(\lambda)$.

2.1.2 Surface Descriptors

The goal of colour constancy is to transform the sensor response vector $\underline{p}^x$ (henceforth underscoring denotes vector quantities) to its descriptor $\underline{d}^x$, where $\underline{d}^x$ encodes 3 properties of the surface reflectance and is invariant to $E(\lambda)$. Of course surfaces with different spectral reflectance functions should have different invariant descriptions. Formally:

$$\underline{d}^x = \mathcal{T}(E(\lambda); \underline{p}^x) \tag{2.2}$$

That is $\underline{d}^x$ is a illuminant dependent transformation of $\underline{p}^x$. In a Mondriaan world, under uniform illumination, a single transformation will apply throughout the image. The transform $\mathcal{T}$ is often considered to be linear. In this case $\mathcal{T}(E(\lambda); \underline{p}^x) = \mathcal{T}(E(\lambda))\underline{p}^x$. So if the number of sensor classes is 3, $s = 3$, then $\mathcal{T}(E(\lambda))$ is a 3×3 matrix.

2.1.3 Continuous functions as Vectors

A 1-dimensional function $F(\lambda)$ which varies slowly with respect to λ can, in a closed interval of λ, be approximated by a vector. Thus the functions of lambda introduced in the preceding section can be described by their values at a discrete number of wavelengths over the visible spectrum. We use spectra where λ is sampled at 10nm intervals from 400 thru 650 nm (vectors then have 26 components). Hence we can rewrite the $R(\lambda)$, $C(\lambda)$, $E(\lambda)$ and $S(\lambda)$ in terms of there corresponding vectors: $\underline{R}$, $\underline{C}$, $\underline{E}$ and $\underline{S}$.

Let us group the s sensors in the $26 \times s$ matrix $\mathcal{R}$. The kth column of $\mathcal{R}$ is the kth receptor vector. We can now rewrite the integral of equation (2.1) in terms of summations:

$$p_k^x = \sum_{i=1}^{26} \mathcal{R}_{ik} \underline{C}_i \tag{2.3}$$

Equation (2.3) is exactly the vector dot-product of the kth sensor with the colour-signal. Hence we can calculate the s sensor responses via equation (2.4) (the k index

is dropped):

$$\underline{p}^x = \mathcal{R}^t \underline{C} \tag{2.4}$$

where t denotes matrix transpose. Thus we can think of a sensor response as the projection of a colour signal onto the sensor axis. The vector representation is useful for analysis since it is impossible to measure precisely complete spectral functions. In addition the techniques of vector algebra are employed in many of the computational strategies for colour constancy.

2.1.4 Finite-dimensional Models

Colour vision can be modelled using a finite dimensional linear model for surface reflectance and illuminant spectra[27]. Let $\mathcal{S}$ be a matrix of d^S (dimension of S) refelectance basis vectors; $\mathcal{S}$ is $26 \times d^S$. Thus a surface reflectance vector $\underline{S}$ is approximated as:

$$\underline{S} \approx \mathcal{S}\underline{\sigma} \tag{2.5}$$

where $\underline{\sigma}$ is a d^S component column vector of weights. Maloney[22] presents evidence which suggests surface reflectances can be well modelled by a set of between 3 and 6 basis vectors. Similarly illuminants are often modelled by a small set of basis vectors. Let $\mathcal{E}$ be the matrix of d^E basis vectors, then:

$$\underline{E} \approx \mathcal{E}\underline{\epsilon} \tag{2.6}$$

$\underline{\epsilon}$ is an d^E dimensional vector of weights. Judd[18] measured 605 daylight illuminants and showed they are well modelled by a set of three basis functions. However many artificial illuminants are poorly approximated using this basis. This is especially true for the spiky illuminant spectra generated by fluorescent lighting.

2.2 General Linear Transforms

In section 2.1.1. we introduced the colour constancy transform $\mathcal{T}$. Almost all authors consider $\mathcal{T}$ to be a linear map (i.e. a 3×3 matrix). If the descriptor for a surface $\underline{S}$ is defined to be its response vector under a canonical illuminant $\underline{E}^C$ then Forsyth[10] has shown that $\mathcal{T}$ **must** be linear. In this section we consider colour constancy algorithms which assume a general linear map. Those algorithms which restrict $\mathcal{T}$ to being diagonal are discussed in section 2.3.

2.2.1 Experimental Performance

We begin by considering the theoretical constancy of a linear transform $\mathcal{T}$. If good constancy is possible then this validates Swain's idea of colour constancy preprocessing. Good theoretical constancy performance is also required in developing illuminant invariants—discussed in chapter 3.

To my knowledge, the theoretical bounds on colour constancy have not been published in the colour constancy literature. Previous studies have estimated how well particular algorithms solve for colour constancy. Thus the results presented here, are by themselves, of considerable interest.

We consider two sets of sensor sensitivities: the cone fundamentals derived by Vos and Walraven[33] and a set of camera sensitivities. The camera sensitivities were derived by multiplying the spectral sensitivity function of our CCD camera by the Kodak Wratten filters #25 (red), #58 (green) and #47B (blue). The sensor sets are shown in Figures 2.1 and 2.2.

We use a set of 7 illuminants: 5 Judd daylight spectra [18], CIE standard illuminant A [33] and a black-body radiator with colour temperature 3600K. These illuminants are applied to a set of 40 surface reflectances consisting of 12 ceramic tiles [3], the 24 Macbeth colour checker [25] patches and 4 of the natural surfaces measured by Krinov[19]. Since the Krinov, ceramic and Macbeth spectra are measured in different units, all surface reflectances are normalized—they are scaled such that their squared area is equal to one. This normalization has little effect on the results presented in this section.

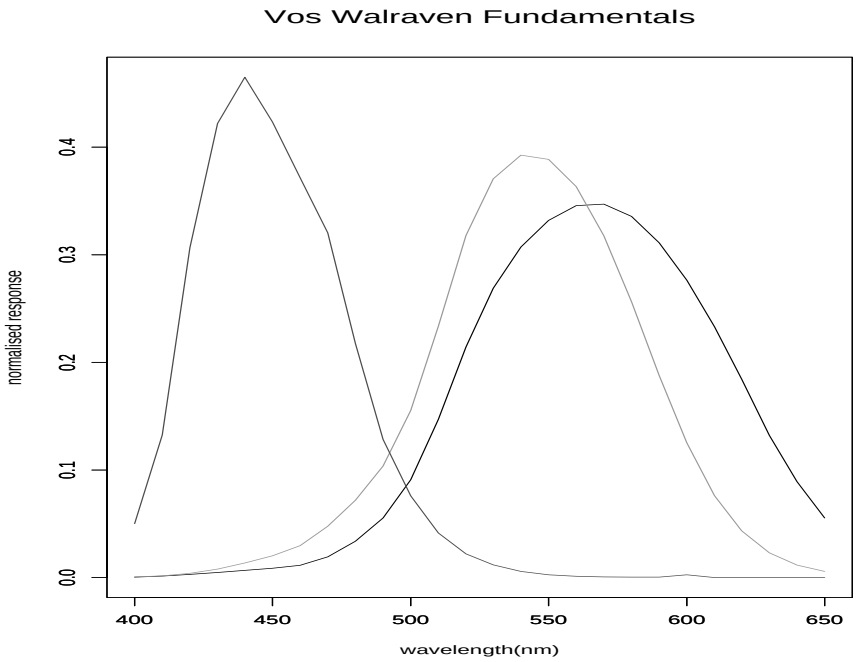

Figure 2.1: Vos Walraven Fundamentals

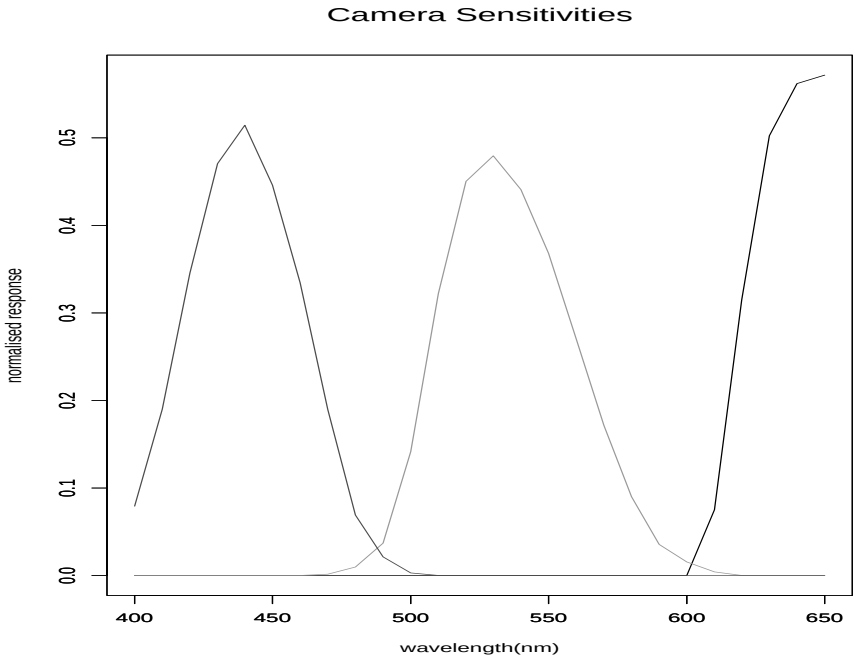

Figure 2.2: Camera Sensitivities

We consider colour constancy to be achieved if response vectors are mapped to their appearance under a canonical illuminant. In these experiments we chose Judd's D55 (55 stands for 5500K) as the canonical illuminant. This implies that descriptors are response vectors for surfaces viewed under D55.

Let $\mathcal{V}$ be a 3×40 matrix of sensor-response vectors generated for the 40 surfaces observed under D55. Similarly, let $\mathcal{W}$ be the matrix of response vectors of the surfaces imaged under another arbitrary illuminant E. To the extent that linear transforms suffice for colour constancy, $\mathcal{V}$ and $\mathcal{W}$ should be approximately equivalent under a matrix transform:

$$\mathcal{V} \approx \mathcal{T}\mathcal{W} \tag{2.7}$$

We solve for the non-zero $\mathcal{T}$ which minimises the sum of the squared error in equation (2.8):

$$minimize \sum_{i,j} ([\mathcal{V}]_{ij} - [\mathcal{T}\mathcal{W}]_{ij})^2 \tag{2.8}$$

The solution for $\mathcal{T}$ is given by the Moore-Penrose inverse $\mathcal{T} = \mathcal{T}\mathcal{V}(\mathcal{W})^+$, where $\mathcal{W}^+ = \mathcal{W}^t[\mathcal{W}\mathcal{W}^t]^{-1}$. Given a fixed set of sensor functions, the solution of equation (2.8) yields the best transformation that takes observations under one illuminant into observations under another.

For both the Vos Walraven and camera sensor sensitivities, we generate experimental data to test how well linear-transform algorithms can possibly perform. For each illuminant, we find the optimal linear transform (by solving equation 2.8) mapping the sensor response vectors for the surface reflectance set to their appearance under the canonical illuminant D55.

Since sensor responses are 3-vectors any metric for evaluating colour constancy should compare fitted vectors (the ith column of $\mathcal{T}\mathcal{W}$) with their corresponding descriptors (the ith column vector $\mathcal{V}$). The euclidean distance between fitted vector and descriptor, normalized with respect to the descriptor length, is a reasonable choice of metric and we will denote this normalized fitted distance as NFD.

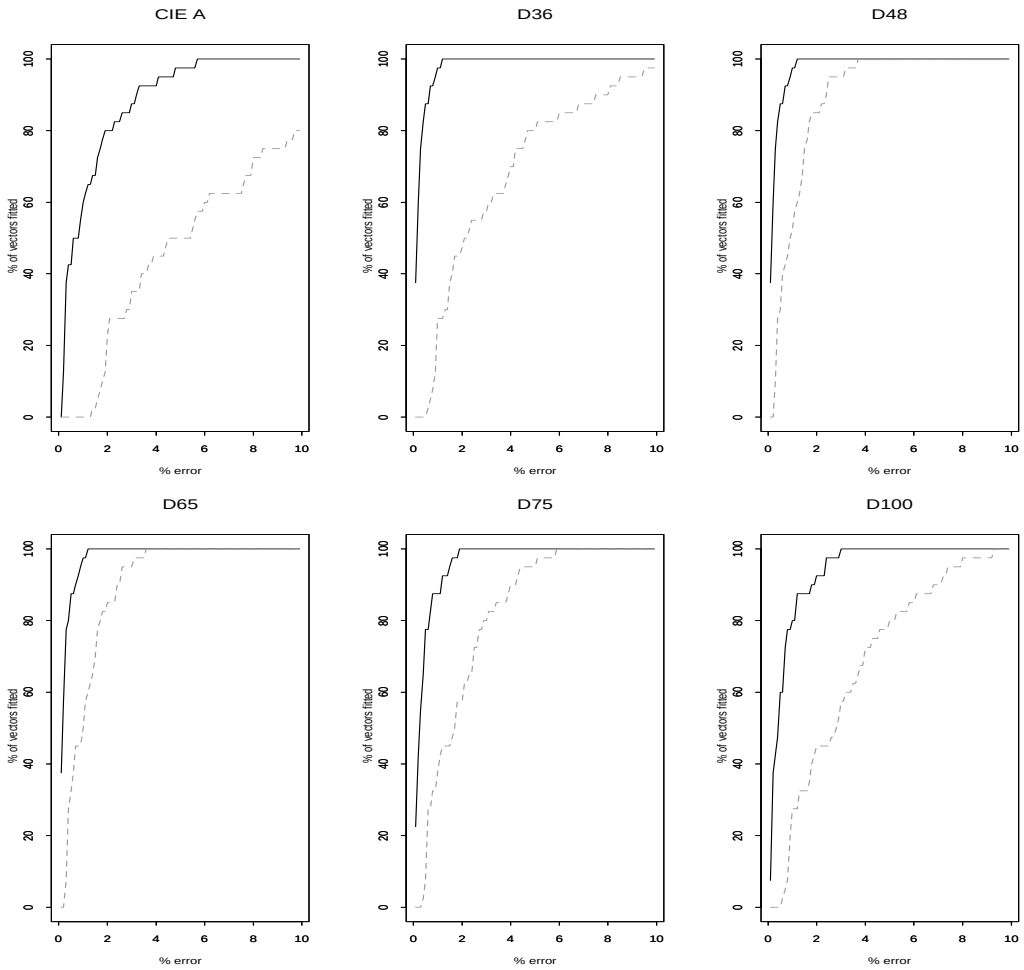

Figure 2.3: Cumulative NFD histogram for Vos Walraven fundamentals

Cumulative NFD histograms for the Vos Walraven and camera sensors are generated for each illuminant, see solid lines in Figures 2.3 and 2.4. In all cases response vectors are mapped to within 10% of their descriptors. These experiments demonstrate that a linear transform is a suitable mechanism for colour constancy. In the following 3 sections we consider computational approaches to finding $\mathcal{T}$.

2.2.2 Gershon's algorithm

Gershon[13] developed an algorithm to solve for $\mathcal{T}$ by making 3 assumptions about the Mondriaan world:

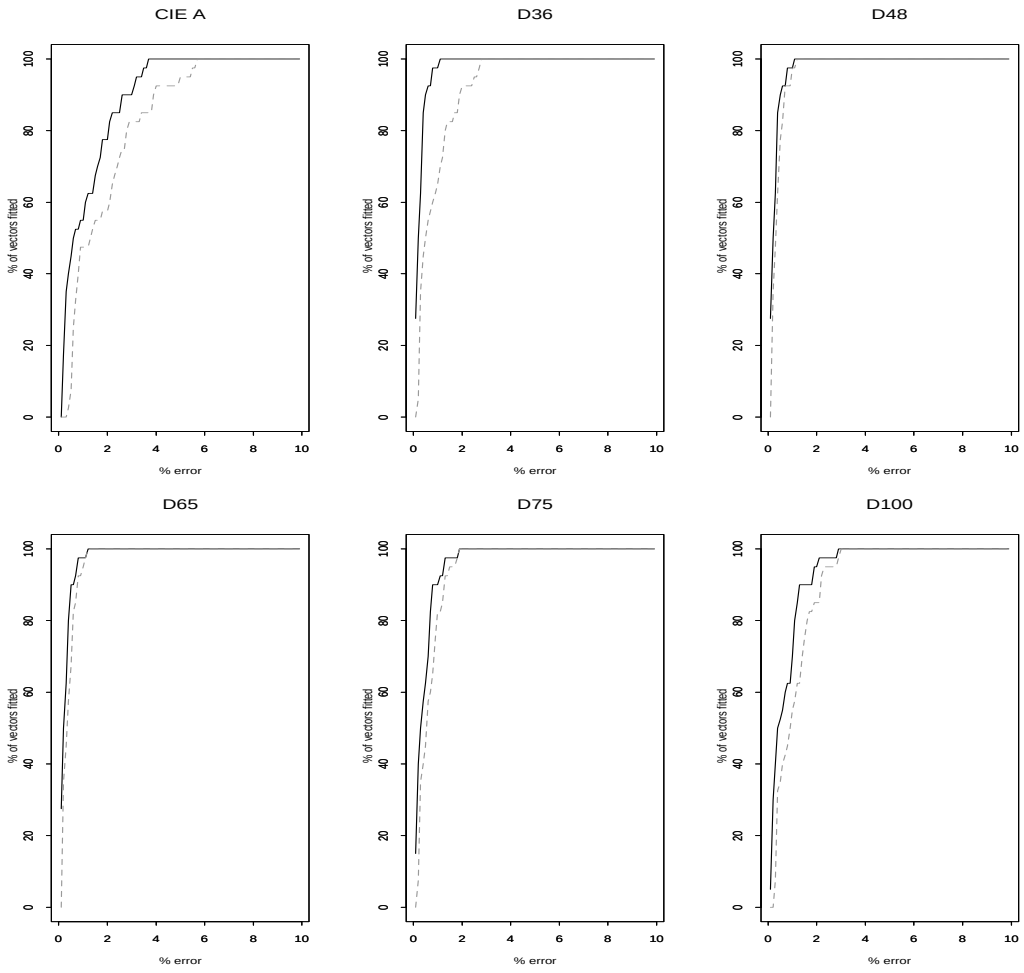

Figure 2.4: Cumulative NFD histogram for Camera Sensors

1. Illumination and surface reflectance spectra are well modelled by small dimensional basis sets. Specifically if there are s sensors then $d^E, d^S \leq s$. We will assume $s = 3$

2. The average of all the distinct surface reflectances in every Mondriaan is the same. We denote the average reflectance vector as $\underline{A}$.

3. Illumination is everywhere constant.

The algorithm proceeds in two stages: firstly assumption 3 is exploited to solve for the illuminant; thereafter $\mathcal{T}$ can be constructed. Let $\Lambda(\underline{A})$ be the 26×3 matrix

constructed by multiplying each column of $\mathcal{E}$ by $\underline{A}$. The illuminant is defined by $\mathcal{E}\underline{\epsilon}$. Thus the average response vector recorded for a Mondriaan can be written as:

$$\underline{p}^{av} = \mathcal{R}^t \Lambda(\underline{A})\underline{\epsilon} \tag{2.9}$$

Since $\mathcal{R}^t\Lambda(\underline{A})$ can be precalculated, ($\Lambda(\underline{A})$ is the same for all Mondriaans) and $\underline{p}^{av}$ can be derived from the image, we can solve for the weight vector $\underline{\epsilon}$:

$$\underline{\epsilon} = [\mathcal{R}^t\Lambda(\underline{A})]^{-1}\underline{p}^{av} \tag{2.10}$$

The illuminant vector is calculated as $\underline{E} = \mathcal{E}\underline{\epsilon}$. Let $\mathcal{Q}$ denote the 26×3 matrix constructed by multiplying each column of $\mathcal{S}$ by $\underline{E}$. The response vector corresponding to a surface reflectance defined by the weights $\underline{\sigma}$ satisfies the following relationship:

$$\underline{p} = \mathcal{R}^t\mathcal{Q}\underline{\sigma} \tag{2.11}$$

Both $\underline{p}$ and $\mathcal{R}^t\mathcal{Q}$ in equation (2.11) are known. Hence we can solve for $\underline{\sigma}$ (the surface descriptor) by calculating:

$$\underline{\sigma} = [\mathcal{R}^t\mathcal{Q}]^{-1}\underline{p} \Rightarrow \mathcal{T} = [\mathcal{R}^t\mathcal{Q}]^{-1} \tag{2.12}$$

For general Mondriaans Gershon's algorithm exhibits poor colour constancy since the average reflectance spectra can vary significantly. Further constraining the illuminant to being everywhere uniform is an unrealistic restriction. Regarding colour indexing, it is highly unlikely that every object will have the same average colour.

2.2.3 Maloney's algorithm

Maloney's algorithm, like Gershon's, proceeds in two stages: firstly the illuminant is estimated, thereafter the constancy transform $\mathcal{T}$ is constructed. However Maloney makes different, weaker, assumptions about the world:

1. If there are s sensors then $d^E \leq s$ and $d^S < s$. We will assume $s = 3$ hence $d^S \leq 2$.

2. Illumination is locally constant.

Given the illuminant vector $\underline{E}$ then $\mathcal{T}$ is calculated by equation (2.12). However, because surface reflectances have dimension 2, $\mathcal{T}$ will be a 2×3 matrix (this also implies the inverse of equation (2.12) is a pseudo-inverse). Thus $\mathcal{T}^{-1}$ is an injective mapping taking 2-dimensional surface weight vectors onto 3-dimensional sensor responses. Alternately a response vector can be thought of as the sum of the 2 columns of $\mathcal{T}^{-1}$ from which it follows that sensor response vectors lie on a plane. Maloney uses this plane constraint to solve for the illuminant.

At this point it is useful to count the number of equations and unknowns. This will lead to a statement about the number of sensor responses needed to solve for $\underline{E}$. Given a single response vector we have 3 knowns and 5 parameters to solve for: $\underline{\epsilon}$ and $\underline{\sigma}$. Adding a second response vector increases the knowns to 6. However, since the illumination is locally constant, the unknowns increase by 2 to 7. By initializing $\underline{\epsilon}_1 = 1$ we reduce the number of unknowns to 6 and hence have enough knowns to solve for the two surface reflectances and the illuminant[1]. Theoretically this implies we can solve for colour descriptors at the edge of two coloured regions.

To solve for $\underline{\epsilon}$ we must first find the normal to the response plane. The normal corresponds to the vector orthogonal to the 2 response vectors and is defined by their vector cross product[2]. We denote the plane normal as π. The responses of any two surfaces, S_1 and S_2, (which are linear combinations of the 2 basis vectors) must lie on the response plane. This implies:

$$\pi^t \mathcal{R}^t \Lambda(\underline{S}_1)\underline{\epsilon} = 0 \tag{2.13}$$

$$\pi^t \mathcal{R}^t \Lambda(\underline{S}_2)\underline{\epsilon} = 0 \tag{2.14}$$

The left-hand sides of equations (2.13) and (2.14) are 3×1 row vectors. Since $\epsilon_1 = 1$ there are exactly 2 equations and 2 unknowns. Thus we can solve for the illuminant; thereafter $\mathcal{T}$ is calculated by equation (2.12).

[1] Setting $\epsilon_1 = 1$ fixes the length of the $\underline{\epsilon}$ vector but does not change its direction.

[2] The plane normal calculated as the cross product of two vectors is susceptible to image noise. More robust estimates of the plane normal can be made by accumulating evidence from many response vectors.

A standard colour vision system, with three sensors, can only achieve colour constancy using Maloney's algorithm if surface reflectances are 2-dimensional. Unfortunately surface reflectances are higher dimensional (between 3 and 6 [22]). Theoretically Maloney's algorithm will perform better if the vision system has more than 3 distinct sensors (there is no published work evaluating this hypothesis). In the case however, more distinct response vectors (of different surfaces) are required to uniquely determine the plane normal π ($n-1$ response vectors are required given n sensor classes). Thus assuming the illumination is only locally constant, Maloney's algorithm can solve the colour constancy problem if there is sufficient, local, colour complexity; where this complexity is defined by the number of sensor classes.

2.2.4 Forsyth's MWEXT

Forsyth develops an algorithm for colour constancy called MWEXT using weak assumptions about the world. In particular surface reflectances and illuminants are not constrained to being finite dimensional. However the illumination is still constrained to being everywhere uniform.

The descriptor for a surface is defined to be its sensor response vector generated under a canonical illuminant. The set of **all** descriptors, $\mathcal{C}$, is used as a constraint in solving for $\mathcal{T}$. All the response vectors in an image must be mapped into $\mathcal{C}$ by $\mathcal{T}$. If $\mathcal{I}$ is the set of image descriptors then:

$$\forall \underline{p} \in \mathcal{I} \quad , \mathcal{T}\underline{p} \in \mathcal{C} \tag{2.15}$$

There may be many linear transforms which satisfy the above constraint; MWEXT (Maloney-Wandell extension) parameterizes the set of candidate transforms. Forsyth suggests that the set of candidates for $\mathcal{T}$ would be diminished by examining other information, for example specularities and mutual illumination. The problem of enumerating the candidate transforms is non-trivial but would certainly be computationally laborious.

Integral to the implementation of MWEXT is the assumption that **all** colours in the world have been seen. Under this assumption the canonical set is a bounded convex

region, or gamut, in receptor space. In this framework, canonical set membership is determined by examining only the boundary, or hull, of the canonical gamut; this significantly reduces computational costs.

The fact that MWEXT returns a set of possible linear transforms, as opposed to a unique answer, has serious implications for Colour-indexing. In particular, MWEXT's multiple solutions suggests that colour constancy is a difficult problem. Thus MWEXT preprocessing will rarely increase the performance of colour indexing with respect to illumination change.

2.3 Diagonal Linear Transforms

Many theories[10, 21, 15] of colour constancy propose that the effect of the illuminant can be discounted by applying a diagonal matrix transform (DMT) to each sensor response vector. In this case colour constancy is achieved by scaling each sensor channel independently. For example the effect of a red illuminant would be discounted by scaling the red sensor catches by a fractional coefficient; thus reducing the magnitude of the red responses.

$$d_x^k = c_k p_x^k \tag{2.16}$$

Recently Forsyth[10] proved that, if surface reflectances are unconstrained, perfect colour constancy can only be achieved via a diagonal matrix transform (and narrow-band sensors. This observation underlines the importance of DMT theories of colour constancy.

2.3.1 Experimental Performance

In this section we consider the theoretical performance of a diagonal transform $\mathcal{D}$. We follow the same experimental procedure discussed in section 2.2.1. In this case we wish to optimize the equality:

$$\mathcal{V} \approx \mathcal{D}\mathcal{W} \tag{2.17}$$

We solve for each row of $\mathcal{V}$ independently, using the Moore-Penrose inverse:

$$\mathcal{D}_{ii} = V_i W_i^t [W_i W_i^t]^{-1} \tag{2.18}$$

For each illuminant we calculate the cumulative NFD histograms for the Vos Walraven and camera sensors, the dotted lines in Figures 2.3 and 2.4. Compared with general linear transform behaviour (solid lines) a diagonal transform, for the Vos Walraven sensors, achieves lower constancy performance. However, a diagonal transform appears an exceptionally good model for the camera sensors. In both cases a diagonal transform achieves good constancy performance.

2.3.2 A Note on the Experimental results

The constancy results for general linear and diagonal transforms, shown in Figures 2.3 and 2.4, are of considerable interest. They bound the performance of all colour constancy algorithms (for the sets of reflectance and illuminant spectra described in 2.2.1). Without considering a particular algorithm we know perfect colour constancy performance is impossible. However, none of the algorithms presented in this chapter formally addresses the question of error in their models. This is a serious short-coming and serves to weaken the applicability of these algorithms.

Also, for the camera sensors, it appears that the best diagonal transform achieves comparable constancy to the best non-diagonal transform. Reconciling this observation with non-diagonal theories of colour constancy would be an interesting line of research.

2.3.3 Sensors and the Diagonal Transform

By examining Figure 2.4, it is clear that the narrow band camera sensors can achieve good colour constancy. In general, narrower sensors imply improved (theoretical) colour constancy. In the limiting case, sensors which are sensitive to single wavelengths can achieve perfect colour constancy. Consider that only the jth component of the kth sensor is non-zero. Then the summation in equation (2.3) can be written as:

$$p_k^x = \mathcal{R}_{jk} S_j E_j \tag{2.19}$$

The effect of the illumination, in the kth channel, is a simple scalar multiplication; which is the same for all surfaces. If all sensors are sensitive to single wavelengths then a diagonal matrix transformation will facilitate colour constancy in an otherwise unrestricted world. In fact since reflectance and illumination spectra tend to vary slowly, a diagonal matrix transformation will work even when the receptors are only relatively narrow band. This explains the constancy success for the camera sensors.

2.4 Von-kries adaptation

One of the earliest models for (human) colour constancy assumes a diagonal matrix transformation. Von Kries [32] hypothesised that chromatic adaptation is a central mechanism for colour constancy. The idea is that over time the eye would adapt to the ambient illumination. Any colour signals are seen relative to this adapted state.

More specifically the Von Kries adapted responses to a surface $S(\lambda)$ in sensor channel k can be written as:

$$d_k^x = \frac{\int S^x(\lambda) E(\lambda) R_k(\lambda) d\lambda}{\int E(\lambda) R_k(\lambda) d\lambda} \tag{2.20}$$

Von Kries conjectures that for any given illuminant E, d_k^x will remain constant. To determine $E(\lambda)$ some authors assume that there is a white (uniform) reflector in every scene. This white patch assumption is common to several algorithms for colour constancy, including Land's retinex theory which is discussed in the next section.

In reality surfaces under varying illuminants are only approximately von Kries invariant. West and Brill[32] derive the conditions surfaces must satisfy for von Kries invariance. Of course all surface reflectances are von Kries invariant if narrow band sensors are employed.

Von Kries invariance is simply a diagonal matrix transformation where the coefficient in each channel is equal to:

$$c_k = \frac{1}{\int E(\lambda) R_k(\lambda) d\lambda} \tag{2.21}$$

2.5 Land's Algorithm

Land's retinex algorithm[20, 21] solves for the coefficients of the diagonal matrix transform by assuming that every scene in the world contains a uniform reflector (with respect to each sensor channel). However, unlike Von Kries invariance, chromatic adaptation is not assumed - that is the eye does not explicitly measure (adapt to) the illuminant. Hence the retinex algorithm addresses the problem of how to relate sensor responses to the white patch. Computation is carried out independently in each sensor channel.

Consider the ratio of the sensor responses at locations x_1 and x_2 i.e. $p_k^{x_1}/p_k^{x_2}$. Clearly if x_1 is fixed then this ratio will be smallest when x_2 corresponds to the white patch (reflectances are between 0 and 1). Further consider a random continuous path which visits $x_1, x_2, x_3 \cdots x_N$. The ratio $p_k^{x_1}/p_k^{x_N}$ can be calculated incrementally as the path is swept out by multiplying local ratios. For example

$$\frac{p_k^{x_4}}{p_k^{x_1}} = \frac{p_k^{x_4}}{p_k^{x_3}} \frac{p_k^{x_3}}{p_k^{x_2}} \frac{p_k^{x_2}}{p_k^{x_1}} \tag{2.22}$$

Land calls a ratio calculated in this incremental manner a *designator*. The retinex algorithm assumes many random paths are generated. At each location the smallest designator value is recorded. If sufficient random paths are generated then the designator at all locations will be relative to the white patch. Thus the illuminant is discounted and colour constant designators derived.

To deal with slowly varying illumination intensity, the local ratios are thresholded. Thus if $p_k^{x_1}/p_k^{x_2}$ is approximately equal to one then the sensor responses are considered due to the same surface reflectance. Accordingly this ratio is set (thresholded) to one.

The white patch assumption is a very strong constraint on the world. Thus Land[21] modifies the retinex algorithm by assuming that the average of all designators at each image location is constant. Brainard[5] shows that if many random

paths of reasonable length are generated then the average designator at x_a is equal to:

$$\frac{(n-1)p_k^{x_a}}{\sum_{i=1}^{N-1} p_k^{x_i}} \tag{2.23}$$

where i indexes all other pixels in the image. Note the illumination terms still cancel and as such that the average designator is constant under changing illumination. However the designator at x will vary as its background changes. For example if the background is predominantly red then $\underline{p}^x$ will be normalized to a red patch changing the background to blue will yield (unsurprisingly) a radically different designator. Brainard[5] demonstrates that less drastic changes in context can significantly alter reflectance designators. We conclude, therefore, that the average designator assumption is at least as strong an assumption as Land's original white patch assumption.

In terms of object identification if an object can be segmented from an image then the average designator will be illuminant invariant and could be used in Swain's algorithm. There are 2 flaws in this reasoning

1. Segmenting an object in an image often requires identifying the object. (This suggests more expensive computation and would not be suitable in an active vision system.)

2. If the object is occluded then the colour designators will change.

2.6 Horn's algorithm

Horn casts Land's retinex in a more rigorous framework. In particular Land's notion of random path is no longer employed. Instead images are normalized to their appearance under a single, uniform illuminant. This allows the responses at any two image locations to be compared directly.

The computational process is summarized below:

1. The logarithm of the colour image (the log-colour image) is calculated; this effectively separates the reflectance and illumination components. Taking the logarithm of both sides of equation (2.16) implies $\log(d_k^x) = \log(c_k) + \log(p_k^x)$.

2. Reflectance changes are distinguished from illuminant variation by examining the Laplacian of the log-colour image. Small Laplacian values are due to illumination gradients; whereas, large values indicate a reflectance edge. Thresholding the Laplacian of the log-colour image effectively removes the spatial variation of the illuminant.

3. Performing the inverse Laplacian gives a new log-colour image. The antilog of this results in an image taken under a single (unknown) illuminant.

Note the above is the essence of the computational process; Horn presents, in detail, the mathematical analysis necessary for its implementation. Horn goes on to suggest a possible biological implementation. However, there are several problems with Horn's algorithm:

1. To solve the inverse Laplacian requires boundary constraints on the Mondriaan (and its image). Namely the Mondriaan must lie completely within an area of constant reflectance. This implies the sensor responses on the boundary of all images must be due to the same surface reflectance.

2. Colour constant descriptors still require a reference patch. Lands white patch or average patch schemes could be used; however, this implies descriptors still depend on the other colours in the scene.

Horn's boundary assumptions are not satisfied even in the simple Mondriaan world. Blake[4] demonstrates that the strong boundary constraints are necessary only because the illuminant component is removed by thresholding the Laplacian of the log-colour image. By applying the threshold on the gradient of the log-colour image Blake develops a computational process which calculates surface lightnesses with weaker boundary constraints.

2.7 Discrete CRULE

All of the colour constancy algorithms, discussed so far, place unrealistic constraints on the world; and as such cannot be used to extend colour-indexing. However we

need not employ a general algorithm for colour constancy. The problem of object identification places constraints on the world; specifically since our goal is to identify an image as being one of a finite set of objects then this implies our world contains a finite, or canonical, set of colours. Thus the colours generated by a colour constancy algorithm must belong to the canonical set.

Solving for the colour constancy transform by enforcing canonical set membership, suggests Forsyth's MWEXT algorithm (2.2.4). However restricting constancy transforms to diagonal matrices leads to Forsyth's second algorithm—CRULE. Like MWEXT, knowledge of all colours in the world is an essential component of CRULE's implementation. In this section we consider a discrete implementation of CRULE.

2.7.1 Colour constraints

The set containing descriptors for all surfaces (of all objects) viewed under the canonical illuminant is called the canonical set, $C = \{\underline{d}^1, \underline{d}^2, \cdots, \underline{d}^m\}$. An arbitrary scene containing ($n \leq m$) distinct surfaces under a single illuminant l generates n image 3-vectors ,$I = \{\underline{p}^1, \underline{p}^2, \cdots, \underline{p}^n\}$. Colour constancy is achieved if we can match each $\underline{p}^i$ to its corresponding canonical descriptor $\underline{d}^j$. The set of all matched canonical descriptors is called a canonical *labelling* and is denoted L (L is a subset of C). Colour constancy can be considered as a mapping $\mathcal{D}$ of I to C, i.e. $\mathcal{D} : I \rightarrow C$.

Since we are assuming a diagonal matrix model of colour constancy then $\underline{d}^j = \mathcal{D}\underline{p}^i$. For each $\underline{p}^i$ there are m possible candidates for $\mathcal{D}$; each image descriptor can be mapped to every member of the canonical set. However each candidate illuminant transform D must map **all** image descriptors onto the canonical set.

2.7.2 The Algorithm

Let the set of possible illuminants (diagonal matrices) which map the ith image response vector onto the canonical set be denoted $\mathcal{D}_i$ (The sets $\mathcal{D}_i$ are easily constructed from equation (2.16)). A diagonal matrix, $\mathcal{D}'$, which maps **all** elements of I into C must be a member of **every** $\mathcal{D}_i$. We can enumerate the set of possible transforms by evaluating:

$$\mathcal{D}' \in \bigcap_{j=1}^{n} \mathcal{D}_j \tag{2.24}$$

Given a valid $\mathcal{D}$ it is straightforward to find the canonical descriptors which correspond to each image descriptor.

A diagonal transform is an approximate model of colour constancy. Hence the intersection of equation (2.24) must deal with the model error; two transforms are considered equivalent if their difference is within model error limits. In section 5, for a set of synthetic Mondriaans and the camera sensors, we use the discrete CRULE as a preprocessing stage for Colour-indexing. Good results are reported.

2.7.3 Discussion

There may be sufficiently restricted identification domains where CRULE can be used. For example consider the object set containing only cereal boxes. If these are always frontally placed with respect to the camera then the discrete CRULE algorithm might work. Unfortunately, for most real sets of objects CRULE algorithm is unlikely to work. A summary of the main reasons for failure (and of the problems to be overcome) are given below:

1. Most objects are 3-dimensional and as such violate the Mondriaan world assumptions.

2. Response vectors can be the result of many confounding processes: including mutual illumination and specularities. These processes acting in local regions of the image have a global effect. Information from all image locations is used as a constraint in CRULE.

3. Background colours may not be in the canonical set.

4. The size of the canonical set can be very large. A larger set implies that there will be a greater number of possible constancy transforms. (As the number of colours becomes large it is pertinent to switch from the discrete CRULE described above to Forsyth's infinite colour CRULE.)

The discrete CRULE can also be used if every image contains a set of reference colours at a known location. In this case the cardinality of $\mathcal{C}$ and $\mathcal{I}$ are the same (the number of reference colours), and hence equation (2.24) returns a unique diagonal transform. Unfortunately, placing known reference colours in every scene imposes a strong constraint on the world; this limits the usefulness of colour indexing.

Swain also suggests using this reference colour constraint. In particular he proposes to solve for the constancy transform with Novak and Shafer's [26] "Supervised colour constancy" algorithm. Unlike CRULE, this algorithm has not been shown to work on real images.

Chapter 3

Robust Object Identification

Swain's Colour-indexing algorithm is remarkably robust to many changes in visual context: including object deformation and occlusion. However, if the colour or intensity of the illuminant changes then Colour-indexing performs poorly. Theoretically the effect of the illuminant can be discounted by applying a colour constancy algorithm to each image. Unfortunately, the colour constancy problem is underconstrained.

To solve for the constancy transform assumptions are made about the world. These assumptions dictate the types of objects which can be identified with Colour-indexing + colour constancy preprocessing. Almost all colour constancy algorithms place strong constraints on the world; these constraints are not satisfied by realistic object sets. Weakening these assumptions, as in Forsyth's MWEXT and CRULE, leads to many candidate constancy transforms. There appears no way to find the correct transform from this candidate set. Further there, is as yet, no colour constancy algorithm which can work in an unconstrained 3-dimensional world. Thus we conclude that colour-indexing cannot reasonably be extended with a colour constancy preprocessing stage.

The failure of colour constancy algorithms leads to a new approach for robust object identification. We propose indexing, not on colours, but rather on illuminant-invariant, or colour constant, image features. The new approach is called *Colour constant colour indexing* and is the major contribution of this thesis. For the rest of this section we assume that a diagonal matrix transform is a reasonable model for

colour constancy. This is clearly true for the camera sensors.

3.1 Opponent Invariants

Hering(1878)[14] proposed that "opponent" combinations of the cone responses are the basis for colour perception. These opponent combinations are named red-green, blue-yellow and white-black (r-g,b-y,w-bl). The idea is that the two colours in each opponent channel compete against each other. For example if the red cone is strongly stimulated and the green cone is weakly stimulated then the r-g opponent channel will give a strong response. The w-bl channel encodes brightness information. Faugeras[9] proposes that the opponent channels are implemented as linear combinations of the logarithm of trichromatic responses:

$$\begin{aligned} r - g = \log(r) - \log(g) \;\Rightarrow\; r - g = \log(r/g) \\ y - b = \log(r) - \log(b) \;\Rightarrow\; y - b = \log(r/b) \\ w - bl = \alpha \log(r) + \beta \log(g) + \gamma \log(b) \end{aligned} \tag{3.1}$$

One of the advantages of this formalism is that the r-g and y-b channels are independent of the **intensity** of the illumination. An intensity change k corresponds to a scalar multiplication of the original trichromatic sensor channels: kr,kg and kb. Hence in the r-g and y-b channels the k component cancels; since multiplication is addition under the logarithm operator. Faugeras hypothesises that, in the biological setting, the the $w-bl$ channel is also illuminant invariant via lateral inhibition between retinal responses. This refers to the idea that the difference of the $w-bl$ at two different retinal locations is independent of the illuminant intensity.

If changing the colour of the illuminant is modelled well by a diagonal matrix transform then all three opponent channels with lateral inhibition are colour constant. To illustrate this consider the $r-g$ channel when the colour of the illuminant changes. The change is modelled by multiplying the r channel by k_1 and the g channel by k_2. The difference in the $r-g$ channel at two retinal locations, a and b, is written as:

$$\begin{aligned} & \log(r_a k_1) - \log(g_a k_2) - (\log(r_b k_1) - \log(g_b k_2)) \\ = \; & \log(r_a) - \log(g_a) - \log(r_b) + \log(g_b) \; = \; \log(\frac{r_a g_b}{r_b g_a}) \end{aligned} \tag{3.2}$$

Faugeras' opponent channels, with lateral inhibition, are invariant to a changing illuminant. Hence this opponent model performs a partial form of colour constancy; colour constant descriptors encode information of one colour relative to another. Nevertheless these illuminant invariant features provide a rich source of colour constant information. As such they are candidate features for object recognition.

Swain[28] investigated an opponent transform for Colour Indexing. However, his opponents are linear combinations of the sensor channels (no logarithms are taken) at unique image locations. In this framework the opponents are not independent of the illuminant. Unsurprisingly Swain's opponent transform does not significantly alter the performance of colour indexing with respect to illuminant change.

3.1.1 Double Opponent Cells

Faugeras proposes that lateral inhibition is implemented as a low frequency attenuating filter. More recently Hurlbert[17] has investigated opponent invariants. Her model is also based on a low frequency attenuating filter—the Laplacian of the Gaussian (LOG). She proposes that this filter is implemented, in humans, by the *double-opponent* cells.

At an early stage of post-retinal computation, area V1 in the visual cortex, *double-opponent* (DO) cells have been identified[16]. These cells have spatially- and chromatically-opponent concentric fields, the centre and surround, each fed by two cone types. For example, the $R{+}G{-}/G{+}R{-}$ cell has a centre which is excited by long-wavelength light and is inhibited by medium-wavelength light. Its surround has an inverse excitation and inhibition.

Hurlbert[17] analyses the operation of the $R{+}G{-}/G{+}R{-}$ cell as:

$$O \; = \; G \star \bigtriangledown^2 \log(\frac{R}{G}) \; = \; G \star \bigtriangledown^2 \log(R) \; - \; G \star \bigtriangledown^2 \log(G) \tag{3.3}$$

G is a Gaussian (smoothing) filter; $\star$ denotes convolution; $\bigtriangledown^2$ is the Laplacian operator and O is its output. Equation (3.3) calculates the Laplacian of $\log(\frac{R}{G})$, $\log(\frac{R}{G}) = \log(R) - \log(G)$, at each point in the visual field. LOG filtering removes the zero frequency component, which in this case is the illuminant, and returns illuminant invariant descriptors. Assuming a spatially varying illuminant the LOG operator must have a small support. The Faugeras y-b and w-bl channels are also illumination invariant under LOG filtering.

3.2 Ratio Invariants

Colour ratios are approximately illuminant invariant; this is implicit in the derivation of the opponent invariants but follows immediately from the diagonal transform model of colour constancy. Consider the ratio of responses in a single colour channel at positions x_1 and x_2:

$$d_k^{x_1} = c_k p_k^{x_1} \quad d_k^{x_2} = c_k p_k^{x_2}$$

$$\Rightarrow \frac{d_k^{x_1}}{d_k^{x_2}} = \frac{p_k^{x_1}}{p_k^{x_2}} \tag{3.4}$$

Ratios of 3-vectors are illumination invariant and form the backbone of colour constant colour indexing. Ratios have a number of favourable properties:

1. ratios can be calculated locally.
2. illumination is only constrained to be locally constant.
3. surfaces in a local neighbourhood will tend to be at similar orientation with respect to viewer and illuminant. Thus ratios will tend to be view point independent.
4. ratios encode spatial **and** colour relationships.

Simple colour ratios also have favourable error properties; especially when compared with the opponent invariants. The question of error is considered in the next two sections.

	% range of fitted relative error		
	red	green	blue
CIE A	[-1.2,1.2]	[-7.3,13.9]	[-3.1,4.4]
36K	[-0.6,0.6]	[-3.5,6.8]	[-0.2,0.7]
D48	[-0.3,0.2]	[-1.1,1.6]	[-1.3,1.3]
D65	[-0.2,0.3]	[-1.8,1.3]	[-1.2,1.3]
D75	[-0.4,0.4]	[-3.1,2.2]	[-2.0,2.1]
D100	[-0.5,0.6]	[-5.3,3.9]	[-3.1,3.3]

Table 3.1: Range fitted relative error in the red, green and blue channels for the camera sensors

3.2.1 Ratios and error

Since a DMT is an approximate model for colour constancy each colour ratio will be constant only within certain error bounds. There is an important connection between relative error of fitted responses and the error in colour ratios which will allow us to use the experiments of chapter 2 to estimate the error in colour ratios.

First we should distinguish between the notions of absolute and relative error. Let us consider the 3-vectors $\underline{p}$ and $\underline{q}$ where $\underline{p} \approx \underline{q}$. There are two methods to determine how closely $\underline{p}$ and $\underline{q}$ match. Absolute error is concerned with the *distance* between $\underline{p}$ and $\underline{q}$. Suitable distance measures (or metrics) include $\sum_{i=1}^{3} |p_i - q_i|$ and $\sum_{i=1}^{3} (p_i - q_i)^2$. The fitting experiments of section 2.3.1 minimise the sum of absolute errors in the best diagonal fit.

However when we compute the ratio of two response vectors we are more interested in the relative error of the result. Relative error compares the ratios of fitted responses $\underline{p}$ and $\underline{q}$ to the unit vector (vector components are divided). Suppose sensor responses lie in the range $[0, 100]$ then the absolute error between responses 0.5 and 0.2 is small. However, the relative error is very large—more than 100%. Relative error between fitted variables can be large only if the variables have small values. In Tables 3.1 and 3.2 we show the maximum fitted relative error for the camera and Vos Walraven sensor sets, for each illuminant.

The maximum relative error shown in Tables 3.1 and 3.2 refers to single surface reflectances and corresponds to the ratio of a fitted 3-vector with its corresponding

	% range of fitted relative error		
	red	green	blue
CIE A	[-19.3,25]	[-17,19.2]	[-6.5,22.6]
36K	[-10.1,12.1]	[-9.1,9.8]	[-1.5,7.9]
48K	[-3.1,4]	[-3.1,2.6]	[-1.9,3.9]
D65	[-3.9,3.3]	[-2.8,3.6]	[-3.7,1.8]
D75	[-6.4,5.8]	[-4.8,6.4]	[-6,3.1]
D100	[-10.2,9.9]	[-8.1,11.3]	[-9.2,4.9]

Table 3.2: Range of fitted relative error in the red, green and blue channels for the Vos Walraven Fundamentals

	% range of ratio error		
	red	green	blue
CIE A	[-2.3,2.4]	[-18.7,22.9]	[-7.1,7.7]
36K	[-1.2,1.2]	[-9.7,10.7]	[-0.9,0.9]
D48	[-0.5,0.5]	[-2.6,2.7]	[-2.53,2.59]
D65	[-0.5,0.5]	[-3.0,3.1]	[-2.5,2.53]
D75	[-0.8,0.8]	[-5.2,5.5]	[-4.1,4.2]
D100	[-1.1,1.2]	[-8.9,9.7]	[-6.2,6.6]

Table 3.3: Range of errors for ratios in the red, green and blue channels for the camera sensors

descriptor. Let p^{x_1} and p^{x_2} denote responses in a single sensor channel corresponding to two surfaces viewed under some illuminant and d^{x_1} and d^{x_2} are their canonical descriptors. Let α be the coefficient which best maps p^{x_1}, p^{x_2} to d^{x_1}, d^{x_2}:

$$\alpha p^{x_1} \approx d^{x_1} \quad , \quad \alpha p^{x_2} \approx d^{x_2} \tag{3.5}$$

Without loss of generality assume $\alpha p^{x_1} = (1+\varepsilon^{x_1})d^{x_1}$ and $\alpha p^{x_2} = (1-\varepsilon^{x_2})d^{x_2}$ where ε^{x_1} is the maximum positive fitted relative error and ε^{x_2} is the minimum negative fitted relative error. We can write the ratio $\frac{p^{x_1}}{p^{x_2}}$ in terms of d^{x_1} and d^{x_2} and thereby make a statement about ratio constancy relative to $\frac{d^{x_1}}{d^{x_2}}$:

$$\frac{d^{x_1}}{d^{x_2}} \approx \frac{(1+\varepsilon^{x_1})d^{x_1}}{(1-\varepsilon^{x_2})d^{x_2}} \tag{3.6}$$

	% range of ratio error		
	red	green	blue
CIE A	[-35.4,54.8]	[-30.4,43.7]	[-23.8,31.2]
36K	[-19.8,24.7]	[-17.3,20.9]	[-8.7,9.5]
48K	[-6.8,7.3]	[-5.6,5.9]	[-5.6,5.9]
D65	[-7.0,7.5]	[-6.2,6.6]	[-5.4,5.7]
D75	[-11.5,13.0]	[-10.6,11.8]	[-8.9,9.7]
D100	[-18.3,22.4]	[-17.5,21.2]	[-13.5,15.5]

Table 3.4: Range of errors for ratios in the red, green and blue channels for the Vos Walraven Fundamentals

It is clear that this ratio has a higher relative error than αp^{x_1}. During a least-squares fit there will be at least one response fitted above (greater than) its descriptor and one response fitted below. Choosing the maximum positive and the maximum negative fitted errors,ε_M and ε_m, we can bound the errors of colour ratios: $[\frac{-\varepsilon_M-\varepsilon_m}{1+\varepsilon_M}, \frac{\varepsilon_M+\varepsilon_m}{1-\varepsilon_m}]$. See Tables 3.3 and 3.4.

3.2.2 Relative error of the Faugeras invariant

From the discussion in the last section it follows that the Faugeras opponent invariant, introduced in equations (3.1), *can* have higher relative error than single channel ratios. Let us consider the $r-g$ channel response at adjacent retinal locations x_1 and x_2. We denote the error in the red and green channels at x_1 as $\varepsilon_r^{x_1}$ and $\varepsilon_g^{x_1}$. Similarly at x_2 the error is $\varepsilon_r^{x_2}$ and $\varepsilon_g^{x_2}$. Rewriting the Faugeras $r-g$ invariant making error terms explicit (where r and g denote descriptors in the red and green channels):

$$\frac{r^{x_1}g^{x_2}}{r^{x_2}g^{x_1}}\frac{(1+\varepsilon_r^{x_1})(1+\varepsilon_g^{x_2})}{(1+\varepsilon_r^{x_2})(1+\varepsilon_g^{x_1})} \tag{3.7}$$

Clearly if $\varepsilon_r^{x_1}$ and $\varepsilon_g^{x_2}$ are both positive and, $\varepsilon_r^{x_2}$ and $\varepsilon_g^{x_1}$ are both negative the Faugeras invariant has a positive error larger than either $\varepsilon_r^{x_1}$ or $\varepsilon_g^{x_2}$. A similar argument holds for an increasing negative error. Clearly the error performance of the Faugeras invariants is linked to the correlation of errors in different sensor classes.

Table 3.5[1] shows the minimum and maximum errors for the $r-g$ and $y-b$

[1] the error data in Tables 3.5 and 3.6 refer to the exponent of the Faugeras channels. This ensures

	% range of opponent error	
	$r-g$	$y-b$
CIE A	[-18.6,22.9]	[-8,8.7]
36K	[-9.7,10.7]	[-1.5,1.6]
48K	[-2.7,2.7]	[-2.6,2.7]
D65	[-3,3.1]	[-2.5,2.6]
D75	[-5.2,5.5]	[-4.1,4.3]
D100	[-8.9,9.8]	[-6.3,6.7]

Table 3.5: Range of error in the $r-g$ and $y-b$ channels for the camera sensors

	% range of opponent error		
	$w-bl$	$r-g$	$y-b$
CIE A	[-32.4,47.8]	[-15.1,17.9]	[-37,58.7]
36K	[-17.7,21.6]	[-8.2,9]	[-20.2,25.2]
48K	[-6.0,6.4]	[-3.1,3.2]	[-7.7,8.3]
D65	[-6.3,-6.37]	[-3.3,3.4]	[-7.86,8.53]
D75	[-10.6,11.8]	[-5.5,5.8]	[-13,14.9]
D100	[-17.0,-20.5]	[-8.9,9.8]	[-20.3,25.5]

Table 3.6: Range of errors for the Faugeras opponent invariants for the Vos Walraven Fundamentals

invariants for the camera sensors. The $y-b$ invariant performs significantly worse than both the red and blue ratios; the $r-g$ invariant performs worse than all three colour ratios. Thus, for the camera sensors, we conclude colour ratios are a more stable index than opponent invariants.

In Table 3.6 the errors in each of the Vos Walraven opponent channels are shown[2]. Clearly both the $y-b$ and $w-bl$ invariants exhibit poorer performance compared to colour ratios. In contrast the $r-g$ channel is less affected by error than *all* the colour ratios and can be considered as a possible index. However taken together the **set** of Faugeras opponents are not a suitable basis for colour constant colour indexing with the Vos Walraven sensors.

a fair comparison to the errors present in simple ratios

[2]The coefficients of α, β and γ defining the $w-bl$ channel of equation (3.1) are set to 0.612, 0.369 and 0.019. This linear combination best matches the performance of the relative luminance efficiency function—$V^*(\lambda)$.

3.3 Volumetric Invariants

Brill[6] develops a theory of colour constancy based on volumetric invariants. Let p_1, p_2, p_3, p_4 denote the response vectors of 4 distinct surfaces viewed under the same illuminant; $\mathcal{M}_{ijk}$ denotes the matrix whose columns are p_i,p_j and p_k. The volume of the parallelepiped bounded by the columns of $\mathcal{M}_{ijk}$ is equal to:

$$v_{ijk} = \text{Det}(\mathcal{M}_{ijk}) \tag{3.8}$$

where Det denotes the determinant function. The volumetric invariant is the ratio of two such volumes: $\frac{v_{ijk}}{v_{ijl}}$. To illustrate lluminant independence consider applying a linear transform $\mathcal{T}$ to the original sensor responses. Under $\mathcal{T}$ the volumetric ratio is written as:

$$\frac{v_{ijk}}{v_{ijl}} = \frac{\text{Det}(\mathcal{T}\mathcal{M}_{ijk})}{\text{Det}(\mathcal{T}\mathcal{M}_{ijl})} = \frac{\text{Det}(\mathcal{T})\text{Det}(\mathcal{M}_{ijk})}{\text{Det}(\mathcal{T})\text{Det}(\mathcal{M}_{ijl})} \tag{3.9}$$

Clearly $\text{Det}(\mathcal{T})$ cancels from top and bottom implying illuminant invariance.

A general linear transform always performs at least as well as a diagonal transform in solving for colour constancy. This is especially true for the extremes in illuminant colour—CIE A and D100. Thus while colour ratios calculated under D55 differ from those calculated under D100, volumetric ratios remain unchanged.

Unfortunately to calculate volumetric ratios there must be at least 4 distinct colours falling in a small neighborhood of the image. Such colour complexity is unlikely; as such we predict that volumetric ratios, used by themselves, would yield poor identification success. However they do provide useful extra information; in particular volumetric ratios encode the interrelationships of 4 surface colours.

3.4 Colour Constant Colour Indexing

All three invariants: ratio, opponent and volumetric are candidate indices for object identification. However, because opponent invariants have poor error properties and volumetric invariants require high colour complexity, we propose indexing only with colour ratios.

Ratios are efficiently calculated in log space via a simple differencing convolution operator. This differencing is, in effect, the derivative of the log-colour image. Unfortunately the first directional derivative is non-isotropic and this could lead to orientation affecting object recognition. Natural choices of isotropic operators include the magnitude of the gradient or the Laplacian. We choose the Laplacian, or more precisely the Laplacian of the gaussian (LOG) so as to include smoothing, because it is simpler to compute and it has a theoretical relationship to the centre surround cells of the human visual system[24].

The LOG operator calculates a weighted average of log differences occurring in a circular field about each image point. Since addition and subtraction in log space corresponds to multiplication and division in non-log space the LOG operator effectively calculates a product of ratios, where each ratio is raised to the power of its weighting coefficient. Each ratio in this product factors out illumination and hence we are assured of the illuminant invariance of the LOG index. Moreover close to the boundary between two coloured regions the LOG operator calculates the weighted product of a single ratio. As such we consider the LOG operator to calculate information similar to explicit ratios.

The simplest Laplacian filter can be written as $(-4_{0,0}, 1_{-1,0}, 1_{0,-1}, 1_{0,1}, 1_{1,0})$, where $-4_{0,0}$ denotes a weight of -4 at mask location $(0,0)$. If i_k^x denotes the logarithm of p_k^x then the Laplacian at image location (x,y) is calculated as $-4i_k^{x,y} + i_k^{x-1,y} + i_k^{x,y-1} + i_k^{x,y+1} + i_k^{x+1,y}$. In non-log space this is equal to:

$$\frac{p_k^{x-1,y}}{p_k^{x,y}} \frac{p_k^{x,y-1}}{p_k^{x,y}} \frac{p_k^{x,y+1}}{p_k^{x,y}} \frac{p_k^{x+1,y}}{p_k^{x,y}}$$

Colour-constant Colour indexing proceeds in three stages:

I. Logarithm step

$$i_k^x \Leftarrow \log(p_k^x) \quad k = 1...3$$

II. Laplacian convolution step

$$d_k^x \Leftarrow \nabla^2 G \star i_k^x \quad k = 1...3$$

III. Histogram step

Steps I and II represent the only additional computation required to obtain illumination independence. The logarithm in Step I can be done by table lookup in hardware and the Laplacian in step II is a separable convolution. As for Hurlbert's opponent invariant, the LOG operator must have a small support so as not to violate the assumption of constant illumination.

We call the histograms of LOG triples *ratio-histograms*. The count in a ratio-histogram bin conveys information, not about colour areas, but about colour boundaries. There are various representational issues resulting from the switch from colours to ratios. These are discussed in Chapter 4. In a number of different experimental conditions colour constant colour indexing performs well. The experiments and results are presented in Chapter 5.

3.5 Colour Constancy by Object Identification

Algorithms which solve the colour constancy problem all make assumptions about the world: retinex assumes that each scene contains a uniform reflector and CRULE has previously seen all surfaces which make up the world. If every world scene contains a model object, at a known location, then colour constant colour-indexing can be used as a preprocessing step for colour constancy. Consider the following algorithm:

1. Focus attention at a known location.
2. Identify object at this location using colour constant colour-indexing.
3. Solve for the constancy transform.
4. Apply this transform to the image thereby generating colour constant descriptors.

Although we do not propose the above as a model for colour constancy, it is interesting to note that Swain requires colour constancy to achieve object identification whereas we can obtain colour constancy as a result of object identification.

The psychophysical experiments of Arend and Reeves and of Craven and Foster, both implicitly address these computational issues in the framework of human vision. We provide a summary of their results in the next section.

3.5.1 Psychophysical Experiments

Machine and human vision share many common goals. As such, studies of the human visual system are often of consequence to machine vision. This is true in the field of active vision—where the goal is to solve specific problems quickly. Colour indexing and colour constant colour indexing both solve the active vision task of identifying a known object (the object is assumed to belong to the set of model objects) at a known location.. Both algorithms address the colour constancy problem. Here we report on psychophysical experiments which partially address identification and colour constancy in the framework of human vision.

Arend and Reeves[1] conducted experiments investigating simultaneous colour constancy. They wished to determine if the human visual system solved the colour constancy problem via simultaneous mechanisms—that is primarily in terms of the spatial interactions among cone responses at different retinal locations, where the eye does not temporally adapt[3]. In their experiments an observer is shown two Mondriaans. The first Mondriaan, the standard, contains n surface reflectances, $S_1(\lambda), S_2(\lambda), \cdots, S_n(\lambda)$, illuminanted under $E^{65}(\lambda)$ (correlated colour temperature of 6500K). The second Mondriaan, the test, is identical to the first except that the incident illuminant is $E^{100}(\lambda)$ (correlated colour temperature of 10000K) and the ith patch, the match, is initialized to $S_i(\lambda)\frac{E^{65}(\lambda)}{E^{100}(\lambda)}$; the ith patch in both Mondriaans reflect **the same colour signal**. (Changing the match reflectance to $S_i(\lambda)$ should render both Mondriaans identical).

Two matching experiments are carried out: chromaticity matching and paper matching. During chromaticity matching the observer is instructed to adjust the chromaticity of the match colour signal such that the ith patch in the test Mondriaan

[3]Land's retinex algorithm is an example of a simultaneous colour constancy algorithm

appears the same as the ith patch in the standard. Colour constancy in this experiment is poor: the observers do not, significantly, alter the chromaticity of the match colour signals. This suggests observers *see* colour signals not surface reflectances.

In the paper-matching experiment the observer is instructed to alter the chromaticity of the match colour signal such that the ith patch in the test Mondriaan looks as if it were cut from the same piece of paper as the corresponding patch in the standard Mondriaan. To aid this matching the observer is encouraged to examine the relationship between colours. Here colour constancy is good.

These matching experiments have two implications:

1. The human vision system does not exhibit simultaneous colour constancy
2. Surface reflectances (or materials) **can** be correctly identified by examining their relationship with other surfaces.

Both these observations favour Colour constant colour indexing. We abandoned colour constancy preprocessing as we judged it to be unattainable. Further colour ratios encode colour relationships between surface reflectances.

Craven and Foster[7] have investigated the problem of operational colour constancy; they pose the question "Can a human observer distinguish between illuminant and reflectance changes?" Their experimental setup is similar to Arend and Reeves. A standard Mondriaan, under a fixed illuminant, is a constant in all experiments. For brief time periods the observer is shown a test which is either:

1. the same Mondriaan under a second illuminant.
2. a Mondriaan with the same spatial pattern but where reflectances are altered.

The observer is asked whether the test and standard Mondriaans differ because of an illuminant or surface change. In all cases observers correctly distinguish between illuminant and surface changes.

This experiment serves to strengthen the work of Arend and Reeves—a changing illuminant **is** identified, and hence colour constancy is not an instantaneous effect.

Moreover the whole test Mondriaan is rapidly matched to the standard suggesting an internal colour constant representation. The LOG of the log colour image, or Faugeras opponent channels, are suitable vehicles for explaining this experiment.

Chapter 4

The Ratio Representation

Switching from colour triples to colour ratios raises several representational issues. Firstly if all colours appear in the world with equal likelihood then this implies the distribution of colour ratios is non-uniform. This is clear from the following simple example. Imagine that colours, in a single sensor class, are integers in the interval $[1, 3]$. Since all colours are equally likely the following ratios will occur with equal probability: $\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{2}{1}, \frac{2}{2}, \frac{2}{3}, \frac{3}{1}, \frac{3}{2}, \frac{3}{3}$. It follows that ratios close to 1 are more likely than ratios close to 3. This simple illustration implies that the ratio histogram should sample ratio space non-uniformly.

We begin this chapter by formalizing the intuition given above. A simple probability model is developed which allows us to solve analytically for the distribution of colour ratios. Thereafter we design a ratio histogram which is optimally sensitive to ratio space: that is, under the assumptions of the model, a randomly generated ratio will fall in each bin with equal probability.

The experiments of section 3.2.1 indicate that colour ratios are only constant (illumination invariant) within certain error bounds. This implies that under two illuminants the same colour ratio may fall in different histogram bins; we call this shifting *ratio migration*. Including ratio error bounds into our probability model allows us to examine ratio migration in detail. In particular, we estimate the probability of ratio migration for different bin distributions.

The chapter concludes by discussing the implications that the nonuniform ratio

distribution and ratio migration have for Colour constant colour indexing.

Note Colour Constant colour indexing indexes not on colour ratios but rather on the LOG of the log-colour image. Clearly these two indices are related. The LOG operator calculates a weighted average of log differences (or ratios). Thus although this chapter deals explicitly with ratios, we expect our results to apply to the LOG index.

4.1 The Probability Model: for Colours

Let us assume that all colours (sensor catches) appear with equal likelihood; that is colours belong to a uniform probability distribution. Formally we write:

$$p_k \in U(1, V) \tag{4.1}$$

which reads, the sensor catch p_k is a random variable belonging to the uniform distribution defined over the interval $[1, V]$. In simpler language an arbitrary sensor catch will have each value between 1 and V with equal probability. The probability that a colour less than X is recorded is equal to:

$$\Pr(p_k < X) \;=\; \frac{X-1}{V-1} \tag{4.2}$$

Since all colours are equally likely each sensor channel must be independent. Therefore the probability of the response vector $\underline{p}$ being less than $(X, Y, Z)^t$ (where corresponding vector components are compared) is calculated as:

$$\Pr(\underline{p} < (X, Y, Z)^t)) \;=\; \frac{(X-1)(Y-1)(Z-1)}{(V-1)^3} \tag{4.3}$$

A cautionary remark should, at the outset, be added to the above model: the assumption of a uniform distribution of colour-vectors is strong and is, in reality, unlikely. Consider the domain of consumer products—cereal boxes etc.—then bright high contrast colours, reds and yellows, are more likely to occurs than browns and mauves. Further the responses in different sensor channels are likely to be correlated. However, Swain's colour histogram uniformly samples colour space; this coupled with

the success of colour indexing makes a uniform colour space a reasonable assumption for analysis.

4.2 The Ratio Distribution

The ratios $\frac{A}{B}$ and $\frac{B}{A}$ contain the same information; hence we define the ratio index to be $\max(\frac{A}{B}, \frac{B}{A})$. Under this definition ratios of sensor catches will also fall in the interval $[1, V]$ but the corresponding probability distribution is non-uniform. What is the probability that $\max(\frac{A}{B}, \frac{B}{A})$ is less than R?

$$\Pr(\max(\frac{A}{B}, \frac{B}{A}) < R) \tag{4.4}$$

Let $R' = \frac{1}{R}$ ($R' \in [\frac{1}{V}, 1]$) then we calculate the probability of 4.4 as:

$$2 * \Pr(\frac{A}{B} > R' \bigwedge A < B) \;=\; 2 * (\frac{1}{2} - \Pr(\frac{A}{B} < R' \bigwedge A < B)) \tag{4.5}$$

$$\Pr(\frac{A}{B} < R' \bigwedge A < B) \;=\; \Pr(A < R'B) \tag{4.6}$$

Assuming that A and B belong to the uniform distribution $U(m, M)$:

$$\Pr(A < R'B) \;=\; \frac{1}{(M-m)^2} \int_m^{MR'} (M - \frac{\alpha}{R'}) d\alpha \tag{4.7}$$

$$\Pr(A < R'B) \;=\; \frac{1}{(M-m)^2} (\frac{M^2 R'}{2} - Mm + \frac{m^2}{2R'}) \tag{4.8}$$

Substituting $M = V$, $m = 1$ and $R' = \frac{1}{R}$ into equation (4.8) and substituting (4.8) into equation (4.5):

$$\Pr(\max(\frac{A}{B}, \frac{B}{A}) < R) = \frac{1}{(V-1)^2} [V^2 + 1 - R - \frac{V^2}{R}] \tag{4.9}$$

We can solve for the density function, $d(R)$, of ratios by differentiating equation (4.9) with respect to R.

$$d(R) \;=\; \frac{1}{(V-1)^2} [\frac{V^2}{R^2} - 1] \tag{4.10}$$

4.3 Optimal Bin Distribution

Equipped with the cumulative ratio distribution, equation (4.9), we can calculate the optimal distribution of histogram bins: ratios should fall in each histogram bin with equal probability. Intuitively this definition of optimality appears reasonable since if most ratios were mapped to a small subset of the histogram bins then this implies different objects would yield similar ratio histograms. However we strengthen this intuition by appealing to the information theoretic notion of *entropy*.

Let the histogram H contain n bins in each dimension (or sensor channel) giving a total of n^3 bins. The event h_{ijk}, that a ratio is mapped to the histogram bin $H(i,j,k)$ occurs with probability P_{ijk}, where $\sum_{i,j,k}^{n} P_{ijk} = 1$. Since we know the distribution of ratios, we can calculate the probability of a ratio sequence and thus a particular histogram. Here we are assuming that we know the number of **distinct** edges which contribute to a particular bin[1] The information contained in a histogram representing m distinct ratios is equal to $m * \text{entropy}(h)$, where h is a random variable defined over the set of events h_{ijk}.

$$\text{entropy}(h) \;=\; \sum_{i,j,k}^{n} -P_{ijk}\log_2(P_{ijk}) \tag{4.11}$$

Entropy is a measure of the average cost of (optimally) encoding each event. Thus $m * \text{entropy}(h)$ is the least number of bits required to encode a histogram with m ratios. If $\forall(i,j,k)\ P_{ijk} = \frac{1}{n^3}$ then entropy(h) is maximum[23], and hence the ratio histogram conveys the most information. Thus an equi-probability partitioning of ratio space is optimal.

Let there be n bins per sensor channel, where the ith bin is sensitive to ratios in the interval $[x_{i-1}, x_i]$ ($x_0 = 1$ and $x_n = V$). For each x_i we must satisfy

$$\Pr(\max(\frac{A}{B}, \frac{B}{A})) \in [1, x_i] \;=\; \frac{i}{n} \tag{4.12}$$

[1] In reality we cannot distinguish between coloured edges with the same ratio triple. However, during histogram matching, this information is partially known. If a histogram bin in one histogram represents 3 edges and has a count of 100 then the corresponding canonical histogram will have a similar bin count in the same bin!.

which implies

$$\frac{1}{(V-1)^2}(V^2+1-x_i-\frac{V^2}{x_i}) \;=\; \frac{i}{n} \tag{4.13}$$

Equation (4.13) can be written as a quadratic in x_i. Thus by finding the roots of equation (4.13) we solve for the bin boundaries x_i.

The above analysis is sufficient for images with only two colours and hence a single edge. If however we introduce a third colour C then the ratios $\frac{A}{B}$ and $\frac{B}{C}$ cannot strictly be considered independent. Incorporating this dependency into the model is non-trivial and will not be considered further here. This dependency disappears if we restrict our attention to the subset of image ratios where all denominators and numerators are **unique** sensor catches.

4.3.1 Optimal Bin Distribution for Camera Sensors

In real images under different illuminants the largest ratio is around 4.5. Assuming that sensor catches fall in the range $[1, 4.5]$ (which could be forced via appropriate scaling) we can solve for the ratio distribution by substituting 4.5 for V in equation (4.9). This distribution is graphed in Figure 4.1. Using equation (4.13) we now solve for the optimal distribution of histogram bins, where like Swain, we divide each ratio channel into 16 bins. The character "O" demarcates bin boundaries in Figure 4.1; the bin distribution is clearly non-uniform.

4.4 Ratio Migration

Sensor ratios are illuminant invariant within certain error bounds. Hence ratios can migrate across bin boundaries as the illumination changes. We wish to incorporate this ratio migration into our probability model. This will allow us to examine the probability of ratio migration for different bin boundaries.

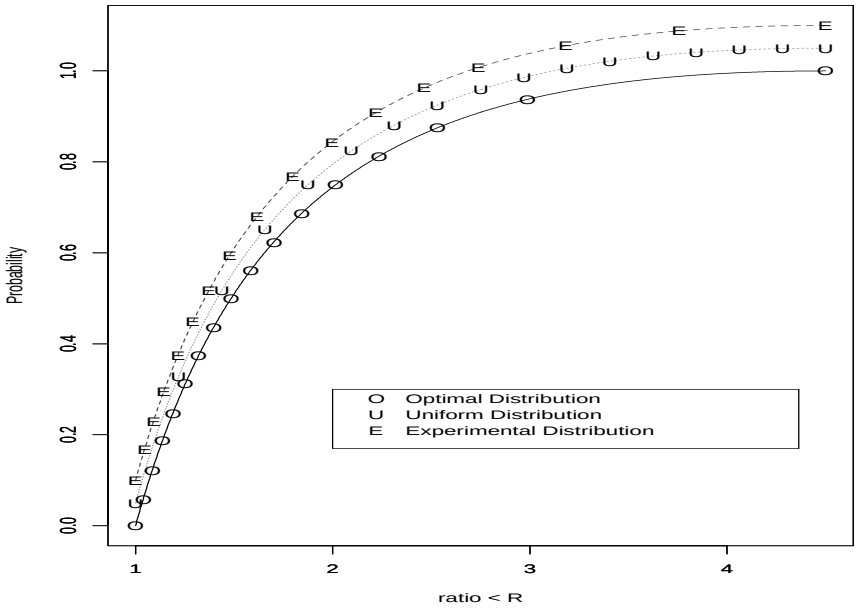

Figure 4.1: Cumulative Probability distribution for colour ratios. Bin boundaries for the optimal, uniform and experimental distributions are denoted by "O","U" and "E".

4.4.1 The Distribution of the Migration Term

In section 3.1.2 we examined the connection between the relative error of fitted responses and the error in colour ratios (fitting refers to the optimal constancy transform which maps sensor values under one illuminant to their appearance under canonical lighting conditions). The relationship between fitted response and standard descriptor is captured below:

$$d_1 \approx (1+\varepsilon)d_1 \tag{4.14}$$

The constancy of equation (4.14) (the variance of the right hand side) is governed by the *migration term* $1+\varepsilon$. Henceforth we assume the migration term of a sensor catch, m^s is a random variable belonging to the unifom distribution $U(m,M)$ ($m < 1$,$M > 1$ and $m, M > 0$. (In reality the migration term will be peaked around one: small errors are more likely than large errors. As such, assuming the migration term belongs to a uniform distribution implies an overestimate of the error. This implies

the following is a worst case analysis).

Given a ratio index $R = \max(\frac{A}{B}, \frac{B}{A})$ we assume the migration term of A (m^A) is independent of the migration term of B (m^B). Hence the combined migration term for R (m^R) falls in the interval $[\frac{m}{M}, \frac{M}{m}]$. We would like to know the probability that m^R is less than equal to some R', $\Pr(m^R < R')$. Assuming $m^R < 1$ then this distribution is given in equation (4.8).

4.4.2 The Probability of Ratio Migration

Given the density function, $d(R)$, of the ratio index, equation (4.10), coupled with the distribution of the ratio migration term allows us to estimate the probability that a ratio will migrate between histogram bins. We split this analysis into two parts: first we determine the probability that a ratio shifts to a higher bin (*positive migration*); thereafter we estimate the probability of a downward shift (*negative migration*).

Let us consider the positive migration of the ith bin (receptive to the interval $[x_{i-1}, x_i]$). The minimum ratio, lying in this interval which can migrate upwards is denoted R_0 and is defined as follows:

$$R_0 \;=\; \begin{cases} x_i \frac{m}{M} & \text{if } x_i \frac{m}{M} > x_{i-1} \\ x_{i-1} & \text{otherwise} \end{cases} \tag{4.15}$$

A ratio R drawn falling in the interval $[R_0, x_i]$ can migrate upwards if and only if:

$$m^R R \geq x_i \;\Rightarrow\; m^R \geq \frac{x_i}{R}$$

Since $\Pr(m^R \geq \frac{x_i}{R}) = \Pr(m^R < \frac{R}{x_i})$ the probability of positive migration in the ith bin is defined by the integral:

$$\Pr(r \in [x_{i-1}, x_i] \bigwedge r m^r > x_i) \;=\; \int_{R_0}^{x_i} d(R) \Pr(m^R < \frac{R}{x_i}) dR \tag{4.16}$$

Let us now calculate the probability of negative migration for the same interval $[x_{i-1}, x_i]$. Firstly we calculate the maximum ratio in this interval which can migrate downwards. This is defined as:

$$R_1 = \begin{cases} x_{i-1}\frac{M}{m} & \text{if } x_i\frac{M}{m} < x_i \\ x_i & \text{otherwise} \end{cases} \tag{4.17}$$

A ratio R falling in the interval $[x_{i-1}, R_1]$ will migrate downward if and only if:

$$m^R R \leq x_{i-1} \;\Rightarrow\; m^R \leq \frac{x_{i-1}}{R}$$

Thus the probability of negative migration in the ith bin is defined by:

$$\Pr(r \in [x_{i-1}, x_i] \bigwedge rm^r < x_{i-1}) = \int_{x_{i-1}}^{R_1} d(R) Pr(m^R < \frac{x_{i-1}}{R}) dR \tag{4.18}$$

Assuming that there is no positive migration from the nth bin and no negative migration from the 1st bin, the total migration probability is calculated as:

$$\Pr(\text{migration}) = \sum_{i=1}^{n-1} \Pr(R \in [x_{i-1}, x_i] \bigwedge Rm^R > x_i) + \sum_{i=2}^{n} \Pr(R \in [x_{i-1}, x_i] \bigwedge Rm^R < x_{i-1}) \tag{4.19}$$

The probability that a sensor catch does not migrate is Pr(no migration) = 1 − Pr(migration). Since the three sensor channels are assumed independent, the probability that a colour ratio vector will **not** migrate, Pr($\underline{R}$ unchanged), is $(1-\Pr(\text{migration}))^3$.

4.5 Experimental Results

Intuitively the analysis of 4.4 implies that the smaller the bin size the greater the effect of ratio migration. Thus a ratio histogram robust to ratio migration should have a small number of large bins. However the optimal bin distribution (4.3) is uneven and includes many small bins— see Figure 4.1. Thus in designing a ratio histogram we need to find a compromise between two conflicting goals:

1. partitioning ratio space into equi-probability regions.

2. minimising the problem of ratio migration.

distribution	*entropy*	Pr($\underline{R}$ unchanged)
optimal	12	0.43
uniform	9.34	0.689
experimental	11.742	0.483

Table 4.1: Entropy versus Ratio Migration

For the camera sensors we calculated 3 different bin distributions: optimal, uniform and experimental. Each is graphed in Figure 4.1. (Note the same cumulative probability distribution is graphed 3 times. For display purposes the uniform and experimental distributions are vertically displaced. This prevents bin boundaries from occluding each other). Like Swain's colour histogram each ratio histogram is partitioned into 16 bins in each sensor channel yielding a total bincount of 4096. The optimal bin distribution refers to the equi-probability partitioning of ratio space defined in equation (4.13). By contrast, the uniform distribution is an equi-volume partitioning of ratio space, where each bin is a cube. The last distribution, the experimental, is implemented in Colour constant colour indexing. (The experimental distribution achieved good match success for all our test images).

Table 4.1 tabulates Pr($\underline{R}$ unchanged) for each distribution, where migration terms are drawn from $[\frac{39}{41}, \frac{41}{39}]$ (this corresponds to a colour constancy fitting error of 2.5% and correlates well with the experiments of 3.2.1). The second column of 4.1 displays the entropy of each bin distribution. This table clearly illustrates the trade-off between discriminatory power (maximising entropy) and robustness to ratio migration. The uniform distribution is the most robust to ratio migration but conveys least information. In contrast the optimal distribution, while maximising entropy, is least resilient to ratio migration. The experimental distribution, implemented in colour constant colour indexing, compromises between entropy and robustness. This compromise is clearly illustrated in Figure 4.1. Where the optimal bin distribution has many small bins—and is especially susceptible to ratio migration—the experimental distribution has fewer bins. In contrast, where the optimal distribution has few large bins the experimental distribution samples ratio space more finely.

The colour constancy fitting error for the Vos Walraven fundamentals is higher than those for the camera sensors. Given fitting errors of 2.5%, 5% and 10%, Pr($\underline{R}$ unchanged) for the experimental distribution is 0.483,0.213 and 0.056. Given the decrease in ratio stability with larger fitting errors, we predict that the Vos Walraven fundamentals will perform poorly for extremes in illuminant colour (CIE A and D100).

4.6 Advantages of the Ratio Representation

Histograming colour ratios has several other advantages:

1. In a world of lambertian surfaces with point source illumination, the ratio of response vectors, corresponding to two surfaces at the same orientation, is viewpoint independent. This is easily demonstrated. Let $\underline{v}$ denote the illuminant direction and $\underline{n}$ the surface normal. The the magnitude of the denominator and numerator responses is proportional to $\underline{v}.\underline{n}$ (the vector dot-product). This term clearly cancels under the ratio operation.

2. Ratios provide more information than colours because a single surface can contribute to many ratios. Consider a $n \times n$ grid of coloured patches. If each colour is unique then the grid contains n^2 distinct colours. However, counting only horizontal and vertical edges, there are $2(n^2 - n)$ colour ratio vectors; all of which can be distinct.

3. Ratio histograms, compared to colour histograms, are less affected by changes in view depth (the distance between object and camera). Since the count in a ratio-histogram bin is a measure of edge length and it is linearly dependent on view depth. This compares favourably with a colour histogram bin which measures area and is therefore proportional to the view depth squared.

4.7 Ratios and Histogram Intersection

Each bin in the ratio histogram is a measure of the length of a particular colour boundary. However the first histogram bin, $H(1,1,1)$, is sensitive to *trivial* colour

ratios—those close to $(1,1,1)^t$. Trivial colour ratios correspond to regions where the colour stays the same and are therefore a measure of area. This implies that two small objects, seen on a large uniform background, will have a large intersection (in both cases there are many trivial ratio vectors). This problem is prevented by removing the first bin from the ratio histogram.

Ignoring trivial ratios implies different ratio histograms will have different total bin counts. Thus care must be taken when normalizing histogram match values. Swain normalizes to the total count in the model histogram. This normalization, for ratio histograms, can result in highly colourful images being falsely matched to a less colourful model. A similar problem occurs if we normalize to the total count in the image histogram. Hence we choose to normalize on the maximum of the image and model bin counts. This ensures a good match occurs only when the intersection is large and both histograms are of similar size. Of course if we presegmented the model images and removed information contributed at an objects boundary we could resort to Swain's model normalization.

4.8 Ratios and 3D geometry

So far we have assumed that the numerator and denominator responses of each colour ratio are drawn from surfaces with the same orientation i.e. we have ignored ratios which occur in tandem with a changing surface normal. Consider two surfaces with normals $\underline{n}_1$ and $\underline{n}_2$, where the light is in direction $\underline{v}$. The corresponding sensor response vectors are written as $(\underline{n}_1.\underline{v})\underline{p}_1$ and $(\underline{n}_2.\underline{v})\underline{p}_2$. Their ratio vector is equal to:

$$\frac{(\underline{n}_1.\underline{v})\underline{p}_1}{(\underline{n}_2.\underline{v})\underline{p}_2} \tag{4.20}$$

Since $\frac{\underline{n}_1.\underline{v}}{\underline{n}_2.\underline{v}}$ is a scalar, the ratio vector normalized to unit length is invariant to the underlying 3D geometry. Under this normalization the ratio histogram encodes 2-dimensional information.

In real objects sharp changes in surface normal often do not coincide with sharp changes in surface colour. For example, close to an orientation boundary, the front

and side of a cereal packet may be the same colour. In this case only trivial ratios $(1,1,1)^t$ will be scaled. If the objects in our database have many orientation edges then the trivial axis, $H(i,i,i)$, can dominate histogram intersection. In this case the trivial axis should be removed.

Chapter 5

Test Results

The colour constant colour indexing algorithm performs well on a variety of real and synthetic images. Objects are correctly identified despite substantial changes in the spectral power distribution of the illuminant. Unsurprisingly, Swain's colour-indexing performs poorly when the illumination changes. It should be noted that in the tests of colour-indexing we use RGB histograms, not opponent-colour histograms (he tests both) and prior background segmentation is not performed on the model images.

To evaluate colour constant colour indexing we first consider whether or not ratios suffice for Swain's original problem under controlled illumination. Second, on synthetic images for which the surface reflectances, illuminants and camera parameters can be completely controlled, we test how the two methods compare. Using these images we go on to evaluate the performance of colour-indexing + colour constancy preprocessing. Finally, we test both methods on real images.

5.1 Tests of the Ratio Representation

Even if colour ratios are independent of illumination, this says little about ratios as a representation for colour indexing. Are ratio histograms sufficient to discriminate between a large number of objects?

To answer this question, we ran the colour constant colour indexing algorithm on

the database of images Swain[1] used in his experiments. First, however, we eliminated 11 of Swain's 66 model images having saturated responses, because ratios relative to saturated pixels cannot be expected to be constant. For our test, then, the model database contains 55 histograms and a second set of 24 different images of the same objects is matched against this database.

Each algorithm's match performance is assessed with reference to three indicators: match rankings, percentile match and match tolerance. The position of the correct match in the sorted list of match values is called its rank, so an image is correctly identified if it has rank 1. The match percentile for each image is defined as $\frac{N-r}{N-1}$, where r is a rank and N is the number of models. Each image is also matched with a certain tolerance relative to the next best matching model. If the correct match has rank i then the match tolerance is $m_i - m_{i-1}$, where m denotes match value. An algorithm that correctly identifies images most of the time, but with high average tolerance, may be preferable to one that correctly identifies images more often, but with lower average tolerance. For each experiment we also calculated the variance of the tolerances. In all cases the variance is small with respect to the average value.

Table 5.1 illustrates the match performance for four algorithms. Swain's, ours with some Gaussian smoothing (LOG indexing), ours with no smoothing (simple Laplacian indexing) and ours where we histogram explicit ratios. Firstly, as Swain reports, colour-indexing works well. The second algorithm, colour constant indexing with smoothing, shows reasonable performance—19 of the 24 images have 1st place rankings. However, match tolerance is much reduced and, more importantly, two of the images are very poorly matched—ranks of 18 and 27.

The poorer performance can in large part be attributed to the effects of too much smoothing. Swain used reduced images of resolution 128×90, which is quite small relative to the 9×9 Laplacian of Gaussian mask. Under these circumstances, colour boundaries will not necessarily be examined in isolation, since the Laplacian operator may straddle more than one edge at at time. To circumvent this problem, we evaluated two further index sets: simple Laplacian filtering (no Gaussian smoothing) and vectors of explicit ratios.

[1]The author is grateful to Michael Swain for providing his images.

Algorithm	*1st Rank*	*Other ranks*	*Av. Perc.*	*Av. Tol.*	*Var. Tol.*
Colour Indexing	23	2	0.999	0.1212	0.005
LOG Indexing	19	3,5,18,27	0.961	0.0613	0.004
Simple Laplacian	21	2,3	0.997	0.0986	0.004
Explicit Ratios	22	2	0.998	0.1023	0.005

Table 5.1: Algorithm Performance : Swain's Images

Histograms of the simple Laplacian of the log-colour image yield the results shown in the third row of Table 5.1. We conclude that the simple Laplacian provides a rich representation for colour constant colour indexing since performance is similar to Swain's colour-indexing.

For the ratio test adjacent pixels in 8 directions are ratioed. Of course, this ratioing can be implemented by a series of directional first-derivative convolutions on the log-colour images. The performance for explicit ratios is similar to that achieved with the simple Laplacian index—see the last row of Table 5.1.

5.2 Tests on Synthetic Images

To the extent that changes in the spectral power distribution of the illumination are modelled by a single scalar multiplication in each sensor channel, the ratio histograms should be relatively illumination independent. To test whether the coefficient rule approximation holds sufficiently for colour ratio indexing, we constructed synthetic images using the measured spectra described in 2.2.1. These images are free from noise, specularities and other confounding processes that could confuse object identification. As such, they represent a minimal world for object identification.

Thirty synthetic Mondriaan objects were generated. Each Mondriaan has the same overall size but contains between 4 and 10 (randomly selected) surface reflectances. If a Mondriaan has m patches, then these are distributed according to the formula: patches in x direction $= \lceil \sqrt{m} \rceil$ and patches in y direction $= \lceil \frac{m}{\sqrt{m}} \rceil$. Patches are, as far as possible, of uniform size. For example if $m = 7$ then the Mondriaan has 3 patches in the first row, 3 in the second, and 1 in the third.

For each illuminant, images of the 30 Mondriaans were generated. To separate the

Algorithm	*No. 1st Rank*	*Failures*	*Av. Perc.*	*Av. Tol.*	*Var. Tol.*
Swain Colour Indexing	20	155	N/A	N/A	N/A
Colour Constant Indexing	180	0	1.000	.568997	0.039

Table 5.2: Algorithm performance : Synthetic images

issue of brightness change from that of hue change in the illumination, the illuminant spectra were normalized such that their squared area is one. Without loss of generality, the Mondriaans imaged under D55 are used as the model set. Match results for Swain's algorithm and for colour constant colour indexing are given in Table 5.2. Note the second column displays the number of match failures. An algorithm fails to identify an image if the intersection with the correct model is zero. If this is the case the match rank is undefined.

As expected, Swain's algorithm performs badly—155 of the 180 Mondriaans have a zero intersection with the correct model. Indeed, colour indexing performs so badly that it is not meaningful to discuss average percentile match or average tolerance. The need for some from of colour constancy is readily apparent.

Colour constant colour indexing performs extremely well. All 180 Mondriaans are correctly identified and with high tolerances.

5.2.1 Biological Plausibility

Using the Vos and Walraven[33] estimate of human cone sensitivities as sensors, we can generate synthetic images and examine to what extent colour constant colour indexing is affected by the choice of cones as sensors. In Table 5.3 we present the theoretical performance results using the cones. The first row contains the match statistics for all 6 test illuminants—i.e. 180 Mondriaans (set1). The second row contains statistics for the test illuminants excluding CIE A and D100 (set2)—120 Mondriaans. CIE A and D100 represent the extremes in the spectral variation of the illuminants.

A comparison of Tables 5.2 and 5.3 reveals that the broad-band nature of the cones does impair the algorithm's performance, but not by too much. Match performance is increased when CIE A and D100—the two extremes of the spectral variation in the illumination—are factored out. Lower rankings result and both the average match

Images	*No. 1st Rank*	*Other Ranks*	*Failures*	*Av. Perc.*	*Av. Tol.*	*Var Tol.*
Set1	135	2-5,7,9,10,14,21	10	0.97	0.194	0.033
Set2	108	2,3,4,5,9,10	3	0.99	0.256	0.034

Table 5.3: Human Cone Performance

Sensors	*No. 1st Rank*	*Other Ranks*	*Failures*	*Av. Perc.*	*Av. Tol.*	*Var. Tol.*
camera	180		0	1.00	0.416275	0.019
Vos Walraven	158	2,3,4,6,11,12,14	3	0.998	0.317	0.025

Table 5.4: Performance for colour-indexing + colour constancy preprocessing.

tolerance and average percentile match increase.

5.2.2 Colour Constancy Preprocessing

The total number of colours appearing in all the Mondriaans is small—exactly 40. Further each Mondriaan has a simple geometric shape. Thus, from our discussion of 2.7. the image of a Mondriaan taken under an arbitrary (but spatially constant) illuminant can be transformed to its appearance relative to a canonical light. That is, we can solve the colour constancy problem for the Mondriaan object set.

We experimentally examine the match performance of colour-indexing + colour constancy (CRULE) preprocessing. The images are created as before: there are 30 canonical models and 180 test images. The match statistics for the camera and Vos Walraven sensors are shown in Table 5.4.

With respect to the camera sensors, colour constancy preprocessing has a dramatic impact on the performance of colour-indexing. Like colour constant colour-indexing, **all** images are now correctly identified (this is in stark contrast to the figures of Table 5.2). However on closer inspection of the data we see that colour constant colour-indexing matches objects with higher average tolerance—0.569 as opposed to 0.416. This disparity suggests that ratio histograms convey more information than colour histograms. Indeed this was predicted in 4.6.

Match performance with respect to the Vos Walraven sensors compares favourably with that obtained by colour constant colour-indexing (see Table 5.3). More Mondriaans are matched at 1st rank, there are less failures and both the percentile match

Database	*No. 1st Rank*	*Other Ranks*	*Av. Perc.*	*Av. Tol.*	*Var. Tol.*
3600K	21	2	0.995	0.165	0.008
4200K	22		1.000	0.145	0.005
5400K	22		1.000	0.137	0.008

Table 5.5: Real Images with Varying Illumination: Colour Constant Indexing

and average tolerance have increased. This success relative to colour constant colour indexing is probably due to the small number of total colours. Consider that the total number of colours were much larger. This implies there will be many candidate transforms which can map image colours onto the canonical set. Since there is no effective means for choosing the correct transform, many false transforms will be chosen adversely affecting match performance. This contrasts with colour constant colour-indexing whose performance is independent of the total number of colours.

5.3 Tests on Real Images

Under three different colour temperatures (3600K,4200K and 5400K) pictures were taken of 11 objects comprised of 3 T-shirts, 3 cereal/detergent boxes, 3 sweaters, a Sun User's manual and a child's toy, for a total of 33 images. When the illumination was changed, so were other viewing conditions; shirts and sweaters were deformed, objects were rotated and occluded. The camera responds linearly with intensity and its spectral response functions are as plotted in Figure 2.2.

Table 5.5 summarizes the match statistics for colour constant colour indexing. A model database was constructed using the 11 images taken under one illuminant and then the other 22 images were matched against it. This was repeated for each illuminant. In the table, each row corresponds to a different choice of model database. Performance is good and is independent of the illuminant.

Table 5.6 tabulates the results for Swain's algorithm. While its performance is poor under varying illumination, it is better than it might have been. This is partly due to the experimental conditions under which the pictures were taken. The colour temperature of the illuminant was changed by placing filters in front of the light

Database	*No. 1st Rank*	*Other Ranks*	*Av. Perc.*	*Av. Tol.*	*Var. Tol.*
3600K	14	2,5,7	0.90	0.08	0.008
4200K	10	2,3,6,7,8,11	0.768	0.066	0.007
5400K	10	2,3,4,8,11	0.80	0.071	0.008

Table 5.6: Real Images with Varying Illumination: Swain's Algorithm

source. Unfortunately, these filters also diminished the intensity of the light. To compensate for this, camera gain and aperture were adjusted. All pictures were made to have pixels which lie close to the maximum camera response (i.e. 255). Both aperture and gain adjustments are linear so should not affect ratio constancy.

Normalizing images in this way encourages Swain's algorithm to work, since these camera adjustments create an approximate form of colour constancy. Nonetheless, even under these favourable experimental conditions Swain's algorithm performs badly. The optimal choice of model set appears to correspond to the 3600K illumination. However, even here 36% of images are wrongly identified; this is extremely poor performance given the small database size. Furthermore, a ranking of 5 or 7 out of 11 is clearly unacceptable.

5.4 Histogram Intersection as a Metric

Swain demonstrates that if two histograms are of the same size then their intersection is a distance metric. In particular, histogram intersection is equivalent to the scaled sum of absolute differences, commonly referred to as the *city-block* metric. Consider the intersection of two histograms M and I each with n bins.

$$if \quad \sum_{i=1}^{n} M_i \; = \sum_{i=1}^{n} I_i \; = \; T$$

$$\Rightarrow \; 1 - (I \bigcap M) \; = \; \frac{1}{2T} \sum_{i=1}^{n} |I_i - M_i|$$

Since ratios close to one are ignored ratio histograms must be normalized to have equal total bin counts. Results for colour constant colour indexing when this metric condition is enforced are given in Table 5.7.

Database	*No. 1st Rank*	*Other Ranks*	*Av. Perc.*	*Av. Tol.*	*Var. Tol.*
Swain's Images	22	5,8	0.992	0.098	0.005
Synthetic Images	180		1.00	0.558	0.034
Real Images	19	2	0.986	0.11	0.004

Table 5.7: Matching when Histogram Intersection is a Metric

It appears that colour constant colour indexing continues to work well. There is a slight performance fall; however, this may not be surprising considering the type of normalization. The results in Table 5.7 illustrate the stability of ratio histograms as a context-invariant object descriptor. Since normalized histograms have constant size, this stability suggests match performance could be increased by using a K-nearest neighbour classifier. For a discussion of these classifiers see Duda and Hart[8].

The model database for a K-nearest neighbour classifier contains the ratio histograms of objects imaged in many different visual contexts. An image is identified by examining the K best matches in this duplicate database. Of course, if all K matches are of the same object then this is a strong match. In general, however, it is sufficient to select the most numerous matched object as the identity of an image. A K-nearest neighbour classifier requires that ratio histograms be a stable representation and that their intersection be a metric. We predict, but have yet to test, that the performance of colour constant colour indexing will be improved if a K nearest neighbour classifier is used. Swain's method cannot be extended in this way, since colour histograms are not stable under illumination change.

Chapter 6

Concluding Remarks

The work presented in this thesis can be extended in various ways. A more informed data analysis would lead to an improved bin distribution for ratio histograms; this in turn would lead to improved object identification. Identification success would also be increased if colour areas **and** colour boundaries contributed to match success. We discuss both of these topics in forthcoming sections. Thereafter we consider using ratio histograms for the object location problem. The chapter concludes with a brief summary of the possible applications for colour constant colour-indexing.

6.1 Data Analysis

The probability model introduced in chapter 4 assumes that both colours and constancy fitting error are uniformly distributed. In reality these distributions are not uniform. The probability model can be strengthened by making more informed estimates about the ratio distribution and the variance of ratios over illuminant changes. This can be determined experimentally.

A more accurate probability model would better guide the choice of bin distribution for ratio histogram. This in turn should lead to improved match success.

6.1.1 Cluster Analysis

When histogram intersection is a metric it is reasonable to consider a K-nearest neighbour approach to object identification (see 5.7). In this case the database contains many histograms for each model (corresponding to a single object imaged in different contexts). There are a set of n histograms corresponding to the ith model: $M_i = M_{i1}, M_{i2}, \cdots, M_{in}$. In this framework identification is a majority decision—the most numerous neighbour of the image histogram identifies the object.

A priori to matching, the space of model histograms can be analysed to determine the likelihood of match success. If each model set occupies a distinct region of histogram space then this favours successful identification. In contrast overlapping model regions is indicative of match failures. This type of *cluster analysis* is useful in evaluating different bin distributions. Further it can provide an upper bound on the number of models which can be successfully identified (as the number of models in the database increase, a false match becomes more likely).

6.2 Lexicographic ordering of colours

Ratio histograms and colour histograms encode related but different information (there is no way to transform one into the other). Thus, an identification system which makes use of both representations would yield improved match success. Unfortunately such a system would necessarily be impaired by a changing illuminant—since colour histograms are not illuminant invariant.

However by altering the implementation of histogram intersection colour histograms **can** be matched independent of the illuminant. The invariance of colour ratios follows from the diagonal matrix model of colour constancy:

$$\underline{d}^x = \mathcal{D}\underline{p}^x \tag{6.2}$$

Another implication of equation (6.2) is that the lexicographic ordering of colours under different illuminants is the same. For example:

$$\underline{p} < \underline{q} \Leftarrow p_1 \leq q_1, p_1 \leq q_2, p_2 \leq q_1$$
$$\Rightarrow \mathcal{D}\underline{p} < \mathcal{D}\underline{q} \tag{6.3}$$

Let us consider two histograms H_1 and H_2 corresponding to the same scene imaged under two illuminants. Subject to the ordering of equation (6.3) let the string of non-empty bin counts of H_1 be denoted S_1; where $S_1 = (c_1, c_2, \cdots, c_m)$, c_i is a bincount and m is the total number of distinct image colours. Hence because colour ordering is maintained during an illuminant change $S_2 = S_1$.

In general the *histogram string* of an object viewed in different visual contexts is not invariant. This is especially true when the object is occluded or when the background varies, in these cases the set of image colours will change. However the problem of approximate string matching occurs often in computing science. For example the UNIX "diff" command finds the minimal difference between two text files. Histogram intersection implemented as string matching will allow colour areas and colour ratios to contribute to object identification.

6.3 Object Location: Histogram Back-projection

The object identification task implicitly assumes that there is a single object in the field of view—histogram intersection compares single model histograms with the colours in an image. Thus histogram intersection is a method for identifying an unknown object at a known location. Swain[28] develops a method, called histogram back-projection for solving the inverse task: identifying a known object at an unknown location.

The location problem is solved in two stages. Firstly the colours which are being searched for are highlighted—a highlight image, h, is constructed. Thereafter we locate the object by finding the densest concentration of highlights in h.

Consider we are searching for the object whose model histogram is M in an image whose histogram is I. We wish to highlight those colours in I that are in correspondence with M. The vehicle for this highlighting operation is the ratio histogram, R,

defined below:

$$R_{ijk} = \min(\frac{I_{ijk}}{M_{ijk}}, 1) \tag{6.1}$$

If the image vector $\underline{p}^x$ maps to I_{abc} then $h^x = R_{abc}$. That is the highlight image is brightest where image colours correspond to model colours. Further the largest bright region in h should correspond to the location of the model (all colours in this region will be highlighted). The brightest region is found via mean-filter convolution.

Histogram back-projection should work equally well for ratio histograms. The only difference is in the highlight image where coloured edges as opposed to coloured areas are enhanced. Thus, although the focus of this thesis is object identification, we predict that illuminant invariants provide a useful basis for the object location problem.

In the context of the human visual system could histogram back-projection serve as a mechanism for controlling eye movements? One problem here is that unlike CCD cameras, the human eye samples each scene non-uniformly; the sampling rate is inversely proportional to the distance from the fovea. However Swain has shown[28] that colour histograms are a salient description for objects despite resolution and hence are suitable location cues[30]. A similar study of the effect of resolution on ratios would give insight on the suitability of ratio histograms for guiding visual attention.

6.4 Applications

Swain proposes two applications for colour-indexing; we review each in relation to the results presented in this thesis. Firstly Swain proposes that colour-indexing can be used in automated check-out devices in grocery stores. Clearly from the results in section 5, colour constant colour indexing should generally perform as well with respect to this task—and in unconstrained illumination. However for objects with few colours, for example fruit, colour constant colour-indexing will fail (since boundaries between

different colours are required). For this restricted domain colour constancy preprocessing may be possible and hence colour indexing employed. If the fruit is placed on an unchanging multi-coloured background there is sufficient colour complexity to solve for colour constancy via the discrete CRULE.

Secondly Swain suggests using colour labels in a robotic manufacturing environment. Colour labels would assist a robot in solving both the identification and location problems. We see no reason why colourful labels and colour constant colour-indexing should not perform equally well in this task. Indeed because manufacturing environments are often illuminated both with natural and artificial light, we predict colour constant colour-indexing will out perform colour-indexing.

6.5 Conclusion

Swain's colour-indexing, whilst remarkably robust to many changes in visual context, is extremely sensitive to varying illumination. Theoretically images can be rendered illumination independent by transforming them via a colour constancy algorithm. Unfortunately colour constancy algorithms place strong restrictions on the types of objects and illuminants which inhabit the world. Even for the least restrictive algorithm, the discrete implementation of Forsyth's CRULE, objects must have simple geometries and the incident illumination is constrained to be spatially constant.

Colour constant colour-indexing indexes not on colour triples but on illuminant invariants, circumventing the need for colour constancy preprocessing. There are three types of invariants—ratio, opponent and volumetric—each of which captures local image properties. All three invariants are useful given complex object sets viewed under spatially varying illumination.

Colour constant colour-indexing, using colour ratios, successfully identifies colourful objects independent of the context in which they are viewed. Further objects are identified with high levels of confidence. We conclude therefore that colour images provide a rich source of information for object recognition.

Bibliography

[1] L. Arend and A. Reeves. Simultaneous color constancy. *J. Opt. Soc. Am. A*, 3:1743–1751, 1986.

[2] J. Beck. *Surface Color Perception.* Cornell University Press, 1972.

[3] R.S. Berns and K.H. Petersen. Empirical modelling of systematic spectrophotometric errors. *Color Res. Appl.*, (4):243, 1988.

[4] A. Blake. Boundary conditions for lightness computation in Mondrian world. *Computer Vision, Graphics, and Image Processing*, 32:314–327, 1985.

[5] D.A. Brainard and B.A. Wandell. Analysis of the reinex theory of color vision. *J. Opt. Soc. Am. A*, 36:1651–1661, 1986.

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

[7] B.J. Craven and D.H. Foster. An operational approach to colour constancy, 1992. To appear.

[8] R.O. Duda and P.E. Hart. *Pattern Classification and Scene Analysis.* John Wiley and Sons, 1973.

[9] O.D. Faugeras. Digital color image processing within the framework of a human vision model. *IEEE. Trans. Acoustics , Speech and Signal Processing*, ASSP-27:380–392, 1979.

[10] D. Forsyth. A novel algorithm for color constancy. *Int. J. Comput. Vision*, 5:5–36, 1990.

[11] B.V. Funt and M.S. Drew. Color constancy computation in near–Mondrian scenes using a finite dimensional linear model. In *Computer Vision and Pattern Recognition Proceedings*, pages 544–549. IEEE Computer Society, June 1988.

[12] B.V. Funt and J. Ho. Color from black and white. In *Proceedings of the Second International Conference on Computer Vision, Tarpon Springs Dec 5-8*, pages 2–8. IEEE, 1988. and *Int. J. Computer Vision*, 3:109-117, 1989.

[13] R. Gershon, A.D. Jepson, and J.K. Tsotsos. From $[r, g, b]$ to surface reflectance: Computing color constant descriptors in images. *Perception*, pages 755–758, 1988.

[14] E. Hering. *Outlines of a Theory of the Light Sense.* Cambridge:Harvard, 1964.

[15] B. K. P. Horn. Determining lightness from an image. *Computer Vision, Graphics, and Image Processing*, 3:277–299, 1974.

[16] D.H. Hubel and T.N. Wiesel. Receptive fields ad functional architecture of monkey stirate cortex. *Journal of Physiology (London)*, 195:215–243, 1968.

[17] A.C. Hurlbert. *The Computation of Color (PhD Thesis).* MIT Artificial Intelligence Laboratory, 1989.

[18] D.B. Judd, D.L. MacAdam, and G. Wyszecki. Spectral distribution of typical daylight as a function of correlated color temperature. *J. Opt. Soc. Am.*, 54:1031–1040, August 1964.

[19] E.L. Krinov. Spectral reflectance properties of natural formations. *Technical Translation TT-439, National Research Council of Canada*, 1947.

[20] E.H. Land. The retinex theory of color vision. *Scientific American*, pages 108–129, 1977.

[21] E.H. Land and J.J. McCann. Lightness and retinex theory. *J. Opt. Soc. Amer.*, 61:1–11, 1971.

[22] L.T. Maloney. *Computational Approaches to Color Constancy.* Stanford University Applied Psychology Laboratory, 1985.

[23] M. Mansiripur. *Introduction to Information Theory.* Prentice-Hall, 1987.

[24] D. Marr. *Vision.* Freeman, 1982.

[25] C.S. McCamy, H. Marcus, and J.G. Davidson. A color-rendition chart. *J. App. Photog. Eng.*, pages 95–99, 1976.

[26] C.L. Novak and S.A. Shafer. Supervised color constancy using a color chart. Technical Report CMU-CS-90-140, Carnegie Mellon University School of Computer Science, 1990.

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

[28] M.J. Swain. *Color Indexing (PhD Thesis).* University of Rochester Department of Computer Science, 1990.

[29] M.J. Swain and D.H. Ballard. Color indexing. *Int. J. Comput. Vision*, 7:11–32, 1991.

[30] M.J. Swain and D.H. Ballard. Low resolution cues for guiding saccadic eye movements, 1991. Submitted to CVPR.

[31] A.M. Wallace. A comparison of approaches to high-level image interpretation. *Pattern Recognition*, 21:241–259, 1988.

[32] G. West and M.H. Brill. Necessary and sufficient conditions for von kries chromatic adaption to give colour constancy. *J. Math. Biol.*, 15:249–258, 1982.

[33] G. Wyszecki and W.S. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulas.* Wiley, New York, 2nd edition, 1982.