A Methodology for the Development of Machine Vision Algorithms Through the use of Human Visual Models

A Thesis
Presented to
The Academic Faculty

by

Wayne D. R. Daley

In Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosopy in Mechanical Engineering

Georgia Institute of Technology
May 14, 2004

A Methodology for the Development of Machine Vision Algorithms Through the use of Human Visual Models

Approved:

Dr. Kok-Meng Lee, Chairman

Dr. Bert Bras

Dr. Paul Griffin

Dr. Ted Doll

Dr. Suresh Sitaraman

Date Approved: May 13th, 2004

Dedication

I would like to dedicate this Thesis to all my teachers. Without your efforts, this would not have been possible.

Special recognition has to be made to my immediate family, Ingrid, my lovely wife, friend, and help mate; along with my sons Nathan and Matthieu. Thank you for your support, patience, and inspiration.

To my brother Michael Daley (Mikey) who showed us by example how to deal with adversity.

I also want to recognize my mother Hermine Daley, who set a fine example of hard work and dedication for all her children.

You have all been my teachers, and to all of you I humbly dedicate this work.

Acknowledgements

Let me begin by thanking Dr. Kok-Meng Lee my Thesis advisor for his encouragement and support at all stages of this effort. His energy and enthusiasm never flagged. His efforts were greatly appreciated. Thanks also to the thesis committee for their comments and suggestions in reviewing the manuscript and also for there perseverance in serving all these years.

Many thanks also to the Georgia Tech Research Institute my employer for the past twenty plus years for providing a work environment that has been fun, challenging and exciting. The financial support this past year that allowed me to devote additional time to this work was greatly appreciated and was instrumental in getting me to this point.

I am also greatly indebted to the ATRP (Agricultural Technology Research Program) and its Director Craig Wyvill for his personal support as well as many of the its programs funded through the State of Georgia. Much of the work described herein was supported through these programs.

My colleagues at GTRI (Georgia Tech Research Institute) have also been a source of ideas, support and encouragement through the years. Thanks especially to Doug Britton for his prayers, constant words of encouragement and his interest. Lucy Johnson also provided me much needed assistance in the generation of many of the graphics.

Lastly, a lot of stars have to be aligned for a boy from a country village in Jamaica to get to this place. Much credit go to people too numerous to mention here that never had the opportunities that I have had, but helped to pave the way for my journey. Thanks does not seem like enough, but I would also like to say that as a result of your prayers and other more practical contributions, I have had blessings in abundance.

Table of Contents

List of Tables

List of Figures

List of Symbols and Nomenclature

α Angle of rays with principal axis of lens

α_c Weighting of center response

α_s Surround Response

$\boldsymbol{\beta}$ Convolution kernel

β_i Element of convolution kernel

∇^2 Laplacian operator

ε Parameter used to compute gaussians

γ Parameter used to compute gaussians

γ_c Gamma for camera

σ Spread of gaussians

σ_c Spread of center kernel

σ_s Spread of surround kernel

μ Mean of gaussians

Ω Solid Angle

Φ Energy Flux

B Blue sensor response

C Pooled center response

Ctr Pooled center response

$C(x, y)$ Contrast at a particular image position

$C_{TOT}(x, y)$ Contrast at a particular color image position

c Subscript to identify center

D_i Distances between clusters

e Energy incident on lens

F F-number of lens

F_{rg} Red green opponent response

F_{yb} Yellow blue opponent response

f Lens focal length

G Green sensor response

G_c Amplifier gains on camera

$G(x, y)$ Two dimensional gaussian function

I_i Input images

I_m Image in image plane of camera model

$I(m, n)$ Image kernel

IR Input red image

IG Input green image

IB Input blue image

J_i Pooled responses over an area for example image

L Long wavelength sensor response

L_q Flux per unit area per unit solid angle

M Medium wavelength sensor response

p Background pixel value for sample image

Q Determinants for partial derivatives

q Target pixel value for sample image

R Red sensor response

R_i Responses of different biological mechanisms

$\mathbf{R}$ Set of responses

r_I Object distance

r_O Image distance

r Dimensionless ratio used to define pixel values

S Short wavelength sensor response

Sur Pooled surround response

s Subscript to identify surround

t_{int} Camera integration time

V Voltage

Z General function describing center surround responses

Summary

In many operations the ability of a machine to "see" is what will determine its effectiveness in its particular domain of operation. For example, in a bin picking problem the ability of the sensing system of a robot to determine the position and orientation of the individual parts will ultimately determine the system's success or failure.

Most systems that require this level of sensing, utilize machine vision in which computers are integrated with image acquisition devices to provide the information required for guidance; as would be needed in a feedback loop for example. The development of algorithms that allow these computers to accomplish the image interpretation has turned out to be less than trivial. This is especially true in the area of natural products such as, meat products, fruit or textile; where, because of their natural variability the ability to develop machine vision algorithms to automatically inspect these products reliably has been problematic.

The goal of this thesis is to attempt to determine a methodology for the integration and streamlining of the process of algorithm development so as to be able to more efficiently develop effective and robust algorithms for this class of problems. Humans, are currently still the best available solutions to these problems. This thesis will examine an approach towards the development of machine vision algorithms using the primate visual system as a model.

The approach taken in this work defines three levels of processing for the visual signal these are sensing, ecoding/transfer, and classification. In particular we examine

the processes of encoding/transfer derived from the results of research in the area of human/primate biological visual processing and their representations. We focus on the use of the receptive field mechanisms that are commonly observed in the human visual system and their processing of contrast in the scenes. We also show that features derived from the responses of these mechanisms are useful for image classification.

Algorithms for implementing these operations are developed using the technique and demonstrated. The other aspect of the approach provides for user guidance by allowing an expert to teach the system by identifying things that are of interest in a particular scene. We then demonstrate development of solutions to three inspection problems using the approach.

Chapter 1

Introduction

1.1 Background/Motivation

Machine vision has been applied extensively and successfully in many automated inspection tasks and bring many benefits to the execution of these operations as compared to humans. The major benefits include, consistency along with more accurate quantitative measures such as size shape and position. Additionally, these systems do not get tired and suffer performance degradation as a result. Machine vision systems are more commonly used where the objects of interest are made to definable tolerances. This is not the case however, with natural products, as variability here is now the rule rather than the exception.

With many machine vision successes in the manufacturing arena along with the advancements in the technology (cheaper and more powerful computers, along with more sensitive and lower cost cameras) more interest is being generated in the inspection of natural products. The variability here is usually several orders of magnitude higher than that for manufactured goods [1]. As a result, most solutions today still have humans in the loop as the typical algorithms are not able to handle the natural change that occurs in the products of interest. According to Graves and Batchelor [1]: " New types of algorithms for image processing are needed." We would also like

to add, that what is also needed, are techniques to support the development of these algorithms.

The design of machine vision algorithms is a complex task in many manufacturing applications, particularly for the automated handling and inspection of natural products. Natural products in this context refer to goods and products such as food, apparel, and textile products [2] where surface texture and reflectance cannot be modified or controlled, and are highly variable. The design of sensing systems for machine guidance and quality control inspection of these entities is a significant task in many industries and attempts to automate these activities have been less than optimal (the systems might for example require frequent retraining). In industrial applications, tasks such as quality control (QC), material handling, and machine guidance are still manually intensive.

The use of machine vision in many of these applications has been stymied because a unique empirical approach is required for each application, due to the lack of a systematic approach to guide the development of imaging algorithms. As stated by Zuech [3]:

> "Successful techniques in manufacturing tend to be very specific and often capitalize on 'clever tricks' associated with manipulating the manufacturing environment."

A similar sentiment is expressed by Arathron [4] in his work titled *Map-Seeking Circuits in Visual Cognition* where he states:

> "General-purpose machine vision remains elusive, and this cannot help but spark a longing to reverse-engineer biology's system, which for the

foreseeable future will set the standard of performance."

The challenge is further increased for natural products as the natural variability that occurs has to be accounted for in some way. General techniques for the development of effective solutions have not been forthcoming.

The paucity of solutions have not been only in the area of inspection but also in its impact on the design of intelligent machinery for conducting operations in these production environments as the ability to sense and respond to naturally varying products is of the utmost importance for realizing machines capable of functioning in these domains. Other areas in which these developments might be off interest include image guided surgery, robotic surgical assistance [5], and medical image analysis such as mammograms [6]. Other applications in which these techniques could be useful is in the interpretation of satellite imagery where it might be possible to develop techniques to enhance images for the viewer to assist in interpretation of visual data.

The objective of this thesis is to establish an engineering foundation that can be used by the machine vision algorithm developer to design workable solutions using a methodology; that is both more efficient and tractable than existing heuristic techniques for finding defects in natural products, while demonstrating robustness to the expected natural variations. The hypothesis here is that we can derive useful approaches from what is currently known about the functioning of the human visual system (HVS).

1.2 Past Research and Related Work

In what follows we provide a review of work in this area. We will begin by covering some of the earlier activity on the influence of the knowledge of biological vision on computer vision and then describe some of the developments that have occurred utilizing these concepts.

In early vision theory and computation, Marr [7] viewed vision as a process that produces from images of the external world a description that is useful to the viewer and not cluttered with irrelevant information. The basic question is how does one go from information to knowledge. What we, in effect have is a mapping from one representation to another. In the case of humans the first representation is known (an image on the retina of the eye) the question is what is the output representation? Marr then proposed a representational framework for deriving shape from images as illustrated in Figure 1.1.

Using this technique, Marr was able to generate useful explanations for the processes involved in generating shape. These included models for using stereopsis, directional selectivity, apparent motion and color. Marr and his colleagues were thus able to describe processes to go from an image to a description of an object's shape in 3D space.

Grinsom [8] extended Marr's work by concentrating on the use of only stereo information in determining the full 21/2 - D sketch but other information could be used as mentioned before, these include texture, color, shading, focusing, occluding contours. Marr[7] and Grinsom [8] also identify three levels at which computations

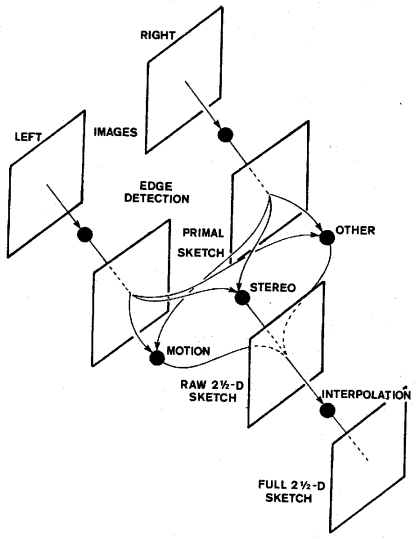

Figure 1.1: Model for obtaining 2 1/2 D sketch

on these image representations can be described , they are: a computational theory, an algorithm to implement the theory and, an implementation of the algorithm. Marr's theory described a process of information or feature extraction from an image that facilitated the determination of shape. The higher level activity for image interpretation was not hypothesized.

Motivated by the area of scientific visualization and the presentation of scientific data, Rogers [9] developed tools to assist the human in the interpretation of visual data by tying the presentation of the data with the cognitive processes that are used to interpret the data. An example of Rogers' applications was that of x-ray radiography [9]. The result of her work was a prototype system called the VIP (visual interaction processor) which provided a link between the visual input image and the problem solving process, the main goal being to handle the functions of hypothesis management and attention direction. She was also able to show that useful results

could also be obtained using this technique to assist radiographers. One aspect of this work which will have direct impact is the development of a process for studying a visual reasoning task and the development of a model for information flow between perception and problem solving along with models for the extraction of information from the people actually doing the task using talk aloud protocols. Specifically this could assist in determining image features of interest based on an expert's description.

Another human performance based approach can be found in the work of Doll et. al. [10] in the development of the Georgia Tech Vision (GTV) model. This model is used to evaluate the conspicuity or detectability of an object in a particular type of background for both still and moving images. The model utilizes findings from computational vision and attention research literature, as well as models based on psychophysical and neurophysiological research. The GTV model utilizes four modules in its implementation; these are called the front end module, the preattentive module, the attentive module and the performance module. The front end module functions like the cells in the retina along with some low-level processing. Preattentive processing simulates perception in the peripheral visual field and identifies objects to be further analyzed in the attentive stage. In the attentive stage close visual inspection or a foveal view is simulated. The processing in the preattentive and attentive phase is implemented as spatial filters the outputs of which drive compressive non-linearity's. The final stage is what is called the performance module in which the probability of locating or missing a given target is computed. The major focus of the work is the prediction of the things that would be conspicuous in a scene as seen by a human observer. This work is significant in that it might be possible to capitalize on this learned behavior of the observer to both identify and appropriately weight

the low-level operations to be performed and also to guide high level interpretation. The other relevant feature is that all the algorithms were developed based on models of the primate visual system. Partial validation of this model has been completed by comparing its performance with that of trained human observers. The results show that the GTV model fairly accurately predicts the probability that objects are located during search and that they are discriminated from background clutter.

Another related work is the use of biological models for early chromatic visual processing by Gershon [11] which resulted in the following contributions:

1. Techniques for determining material changes as opposed to shadow boundaries.

2. Transformations of $[R, G, B]$ to form color constant images (an approach to implement an algorithm for color constancy).

3. The identification of highlights through the use of chromatic information.

Gershon looked at utilizing some of the basic principles from the physiological and psychological knowledge of color vision utilizing information about the mechanisms in color vision whose functions have been determined and documented. Gershon's models follow the structure of the visual pathways; he transforms the inputs into three chromatic components and looked at a linear, a logarithmic and a non-linear adaptation of the outputs from early vision processing. Utilizing these transformations he found reasonable qualitative agreement with responses in human color vision. This implies that an approach to derive computational techniques from biological models might be feasible.

For differentiating between material and shadow boundaries Gershon developed a Relative Amplitude Response (RAR) [11] function as described in Equation (1.1)

$$RAR = \frac{|Peak\ response\ R+G-/R-G+|}{\sqrt{(Peak\ response\ R+/R-)^2+(Peak\ response\ G+/G-)^2}} \tag{1.1}$$

where R and G are the red and green responses; and the '+' and '−' identity the opponent responses. For example, $R+\ G-$ represents the signal obtained by taking the R signal minus the G signal in a region. If $RAR > \frac{|gR-rG|}{R^2+G^2}$ is true where g and r are functions of the ambient illumination and other objects in the scene, then the boundary is due to a material change and not a shadow change. Gershon also identified techniques for simulating color constancy along with algorithms to detect highlights based on the correlation of the reflected spectrum with that of the illuminant.

Direct applications of methods for developing non-linear filters have been demonstrated by Belkacem-Boussaid and Beghdadi [12]. They were able to address the problem of image smoothing while preserving edges. They developed and tested the development of filters that were based on models of the human visual system and showed that they had performance comparable to other established techniques.

Much thought has also been given towards implementing human vision models in hardware. In this work Shah and Levine [13] [14] developed and tested models of human visual performance with the goal of implementing them in silicon thereby imbuing imaging sensors with some of the superior capabilities of the human visual system. They utilized DOG (Difference of Gaussian) filters as representing some of the lower level operations conducted in early vision. This work was done using assuming only achromatic information but they were able to simulate some of the known behaviors on the human visual system.

In order to conduct image segmentation texture differences are many times of

interest. One therefore sometimes has the need to evaluate texture. In this work, Papathomas et. al. [15] develop models for texture segmentation based on human models They develop what they call first order and second order features. The first order texture features are edges based on color or luminance while second order features spatial frequency, binocular disparity and double opponent responses.

Sometimes there is a need to automate the analysis of large quantities of data. The authors here are concerned with the development of algorithms to assist in the analysis of large image sets as would be obtained for example from planetary exploration. In this instance Privitera and Stark [16] are interested in detecting the things humans would detect in analyzing these images. The idea is to develop techniques that would identify features of interest and to use the human visual system as a model in developing these techniques. Among the algorithms suggested is the use of center-surround operators derived from early-stage vision.

In many computer imaging and machine vision tasks the shapes of entities in the scenes are of significant interest. Sakai and Finkel [17] in their work make the point that traditional approaches do not match the human biological system as they are too computationally intensive. They demonstrate approaches using DOG filters in the first stages along with representations in the primary visual cortex.

There are also many other instances in which we utilize knowledge about the human biological system in the design of engineering solutions, audio coding, the telephone system and the television transmission system are common examples. As further motivation, for the approach proposed, we briefly describe the operation of these systems.

The first example to be considered is MP3. This stands for MPEG Audio layer-3.

MPEG stands for Moving Pictures Expert Group and they have guided the development of compression techniques for video and audio. MP3 is the MPEG subsystem to compress sound. It uses a technique called perceptual sound shaping in which it makes use of some of the features of human hearing, for example

- There are certain sounds that the human ear cannot hear
- There are certain sounds that the human ear hears much better than others
- If there are two sounds playing simultaneously we hear the louder one but cannot hear the softer one.

Using this knowledge it is possible to shrink the size of digitally recorded music files by a factor of 10. Thus, the popularity of the MP3 format for the storage and transmission of music files.

Similar information was used in designing the telephone and television transmission system. In designing the phone system the designers used the knowledge that the bandwidth for recognizable speech is about 3kHz. This allows us to design systems that can efficiently utilize the full bandwidth of the medium for transmission of audio. The designers of the system for the transmission of television signals on the other hand utilized the trichromatic theory of color vision in the implementation of that system. It is known from the biology, that we have four kinds of sensors in the eye with three being responsible for viewing scenes in relatively bright situations with the illumination on the order of $10^2 cd/m^2$. Electrically these signals are represented as R,G and B (Red, Green and Blue) to correspond to these sensor responses in the eye. It is known from the biology, however, that we are more sensitive to the green

response than the blue or the red so that we could get by with a high bandwidth signal in the green and a lower bandwidth signal to accommodate the red and blue [18]. This is the approach used in the design of the YIQ signals commonly used in broadcast television. The 1/30th of a second scanning rate of the typical television monitor was also chosen to produce the most realistic signal while utilizing the lowest possible bandwidth for data transmission.

It is clear from these examples that engineering solutions to some specific problems especially as they relate to interaction with humans can be derived utilizing the biological principles under which we function. The overall conclusion from this review is that useful results have been obtained by utilizing human models in many image processing and other domains and the extension of these approaches in the design of machine vision algorithms appears feasible.

1.3 General Problem Description

The literature review above was not meant to be exhaustive but rather to highlight representative work that have a direct bearing on this thesis. While much work has been done in the general computer vision area not much has been done in migrating these developments and knowledge to machine vision solutions. It is observed, however, that many useful techniques have been derived from knowledge of the functions carried out at different levels of the HVS.

The difficulties associated with the development of automated inspection algorithms for these entities are as follows: (1) Natural products are non-uniform; techniques for analyzing natural products must have the ability to accommodate a good

bit of naturally occurring variability. (2) There is usually some subjectiveness [19] involved in the decision making process utilized by the humans who currently conduct these operations. (3) Compounding the problem is fact that inspection standards could also be variable, as in many cases the distribution of the output remains constant no matter the quality distribution of the input; some methodology for judiciously varying the interpretations of the quality standard is therefore a necessity. Combine the above problems with the difficulties inherent in implementing machine vision solutions and many problems quite easily become very challenging.

Solutions are being applied towards several problems in food packaging [20] these include: bottle closure inspection, label inspection, produce grading and sorting, internal contaminant detection and thickness and profile gages. While not natural products directly, they come the closest currently to the solutions that would be of interest. These specific implementations are for problems that are fairly well defined with a clear description of the visual attributes to be identified. For manufactured products the problems are usually more clearly defined and can be specified in terms of flaws that are measurable, errors in dimensions for example. Procedures for tackling these kinds of problems have become common practice (structured lighting, back lighting, subpixel edge detection). Even in this arena, however, the determination of quality parameters have been problematic.

While a variety of skills are required for the successful implementation of machine vision solutions, software development has been one of the more difficult issues [21]. Algorithms with an analytic and computational basis provide at least two advantages for the systems developer: (1) a unified approach to tackling new and unique problems, and (2) the ability to predict performance.

The benefits of the first are obvious and the benefits of the latter lie mostly in the fact that QC systems can be more robustly designed if the performance of the inspector (or inspection system) can be characterized and is consistent. Drury and Fox [22] state that inspection error exerts a significant influence on quality control systems and that inspection QC tasks are error prone with error rates of 25% or higher. It is desirable to be able to accurately measure error rates and to design for them. This can be accomplished by the use of signal detection theory for example. With the current system it is difficult to tell what the true error rates are. Most QC systems, however, assume a perfect inspector and research has shown that it is difficult to characterize human inspection performance as it has been noted that fault detection ability deteriorates as a function of time (40% in 30 minutes [22]). These deficiencies can have a significant economic impact, which manifest themselves in the form of rework, customer dissatisfaction, and machines that don't function according to specifications.

As described before, the standard practice in developing algorithms has been to utilize ad hoc and heuristic techniques. For these reasons, this thesis focuses on the development of a more structured approach towards solving these problems, based on scientific principles, physical measurements, and user guidance.

1.4 Specific Applications

This thesis proposes a more structured and tractable approach towards the solution of quality-control problems in food processing. As illustrative examples, the development and testing of the approach will revolve around two quality control problems

in vegetable and meat processing. Specifically, this thesis will examine fan bone detection on poultry breast fillets and grapefruit grading as shown in Figure 1.2 and Figure 1.3 respectively. Mostly the defects to be identified are not life threatening in nature but rather are visual defects that give the product an undesirable appearance which can result in marketing difficulties for the producer as the product might not be purchased by the consumer or in some cases could be returned to the producer. This could have a significant negative impact on the company bottomline. Currently most of these operations are conducted by people doing visual checks on the line and the concern about the availability of labor to do this job in the future as well as the long term performance of QC inspectors is a driving force for automating these applications.

These two applications will then serve as the testbed for the approach to be described. In these two cases we need to find defects on the products. In the case of the breast fillets we need to find the fan bone located on the middle right of the screen. In most automated systems this is easily confused with shaded regions or blood spots on the product. In the case of the grapefruit we need to determine the amount of discoloration that is present on the surface of the fruit. The severity of the occurrence is determined from the percentage of the area of the fruit that has this discoloration. In the case of the fan bone there is zero tolerance while in the case of the grapefruit there is some acceptable tolerance for the defect.

The human visual system also tends to be more robust than most man-made systems. Using face recognition systems as an example, the performance of some of the more advanced systems on the Feret database [23] were as follows: with frontal images taken on the same day the performance of the systems were 95%. For images

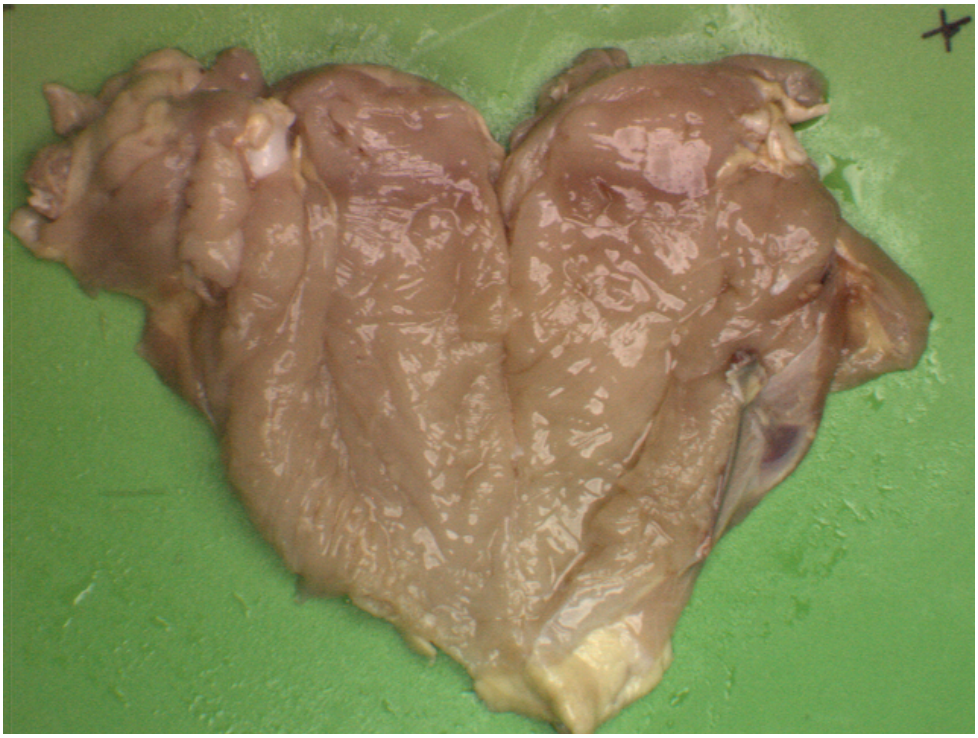

Figure 1.2: Example of a breast fillet with a fanbone

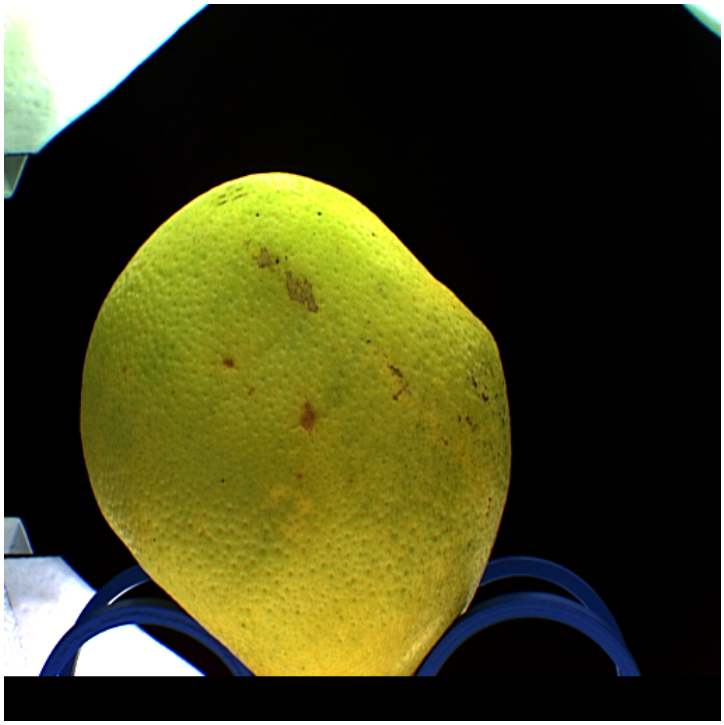

Figure 1.3: Example of a grapefruit with defects

taken with different cameras and lighting the typical performance drops to 80%. With images taken a year later the recognition performance drops even further to about 50% [23]. The main point here is that a human would not typically exhibit this deterioration in performance and there is still a great deal we can learn about the development of practical solutions by studying their biological counterparts.

1.5 Proposed Approach

In this thesis our intention is to extend this paradigm to the design of machine vision algorithms especially as they relate to the subset of problems where people are currently still the best sensors of choice. We will attempt to integrate the knowledge gained from the work in human vision and direct them towards the establishment of an engineering framework for the design of machine vision algorithms. This thesis will then attempt to build on this knowledge to formulate such a framework. The problem domain includes

- imprecise description of defects,
- natural variability in the product, and
- subjectiveness in the interpretation of the data.

The general approach will involve learning from examples, approaches that could be used in solving the problem. This would follow the way things are currently done. If a new employee is hired to function as an inspector he is usually apprenticed to an experienced inspector, this person then usually instructs the apprentice by showing or describing to her examples of the kinds of defects that need to be identified. This

person then learns to recognize them and after some practice period is placed on the production line.

A similar strategy will be employed here where the machine will be shown samples of defective and non-defective product. We will then attempt to derive features based on the processing of the HVS which will then be classified to identify the defects. In this thesis it is planned to develop the approach using two artifacts and to test its performance on both. In one case poultry will be used and grapefruit products the other. These are two example domains in which humans are still the main sensors used for quality and process control. They are, however, representative of a class of problems that today has gone largely unsolved.

1.6 Outline of this Thesis

In the absence of lapses in vigilance or fatigue, the human visual system displays the remarkable ability to function very well, when operating in this somewhat fuzzy domain even in the presence of significant noise. Man-made systems have had much difficulty achieving the same degree of functionality or robustness. The questions to be addressed are as follows: can we effectively utilize some of the proposed human representations and processing techniques along with physical models in man-made systems to improve their performance? Is there a general framework within which the design and development of these systems can be conducted?

The thesis is organized as follows: Chapter 2 will describe the human visual system (HVS) and summarize the state of related knowledge. In Chapter 3 we present a framework for algorithm development. Chapter 4 will describe the models that are

utilized in the investigation. Chapter 5 in turn will describe the testing procedures and the results while the overall results and conclusions will be presented in Chapter 6.

Chapter 2

The Human Visual System (HVS)

The human visual system (HVS) is an amazing machine when one considers its ability to sense and interpret the world around us. It is also a very complicated device all the functions of which are not well understood. One thing that is agreed on, however, is that it allows us to function extremely well in our natural environment. If we are able to inculcate machines with some of these abilities it would allow them to perform in environments that require them to be flexible and to adjust to variability. In this chapter we will summarize some of what is known about the functioning of the HVS.

2.1 Human Eye Structure and Operation

The human visual system could be described as using three sequential processes; sensing, encoding, and transfer. *Sensing* relates to the acquisition of photons. *Encoding* is a data reduction process to allow for *efficient transfer of this information* to the brain, decoding mechanisms exist in the brain to use this data for interpretation of the scene. A description of the process as currently understood will now be presented and is derived from [24], [25] and [26].

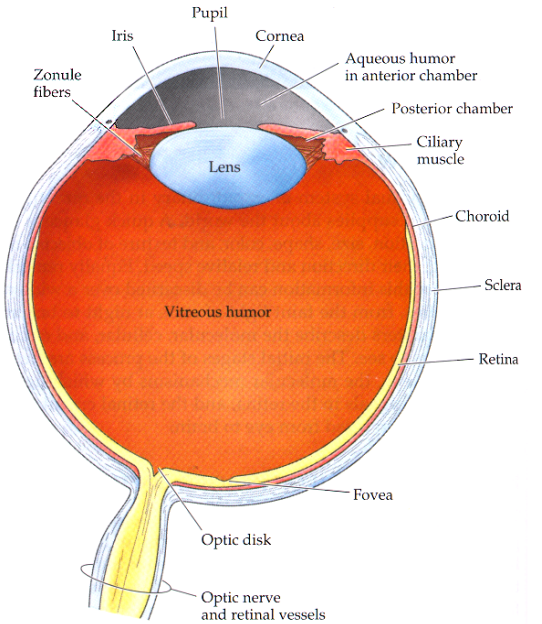

Figure 2.1: Diagram of the eye [25]

2.1.1 Sensing

A diagram of the eye is shown in Figure 2.1, illustrating its principal elements. The eye is approximately a spherical structure about 17 mm in diameter. At one end is an opening that allows light to enter through a lens and focused at the back of the eye on a structure named the retina. In front of the lens is a structure called the iris which controls how much light is allowed to enter the eye. The focal length of the lens is adjustable by muscles attached to the periphery of the lens and the process by which this occurs is called adaptation. The power of a lens is measured in Diopters and is defined as $1/f$ where f is the focal length of the lens measured in meters. The iris can adjust from a diameter of 1.5 to 8.0 mm resulting in f-numbers for the eye from f2.5 to f13 where the f-number is defined as f/d; and d for a human eye is the diameter of the opening of the iris.

The retina houses the sensing elements. Because of the morphology of these sensing elements they are called rods and cones. The rods are long and thin while the cones are relatively more rotund with a conical tip. The rods are primarily responsible for low light level or scotopic vision (< 1 lux) while the cones are dominant at higher light levels or what is called photopic vision. It is said that the rods are capable of detecting one quanta of light after it has been dark adapted for approximately one hour. The retina is also backwards from what would be thought intuitively as the light has to pass through several layers of nervous tissue to get to the sensing elements. This is necessary so that the rods and cones can be replenished with proteins that are depleted as the rods and cones sense the incoming radiation.

The vision process is initiated by light stimulating the photopigments in the rods

and cones in the retina. There are four types of photopigments, one type in the rods, and three distributed among the cones. The photopigments consists of a protein molecule called opsin to which is bound a derivative of vitamin A1. The most studied visual pigment is rhodopsin in the rods and the process to be described is for the rods. A similar process is thought to exist for the cones. When the rod absorbs light a process called cis-trans isomerization takes place in which the outer molecules break away from the opsin, this then generates a small change in electrical potential across the walls of the cell of about $2\mu V$. This voltage change is then transmitted to the neurons connected to the rods and cones. Studies have determined the spectral sensitivity of the four pigments. The rods are centered at $496nm$, the blue (B) cones at $420nm$, the green (G) cones at $530nm$, and the red (R) cones at $560nm$. The visual system actually responds into what should be more accurately termed the short, medium and long wavelength ranges of the visible spectrum or SML; these are typically referred to as RGB. This nomenclature is not accurate, however, and should more appropriately be described in terms of the probability of absorption of photons of light at different wavelengths.

Because of the overlapping spectra of the three cone pigments, there is a unique combination of absorbance probabilities in the visible spectrum; thus, by comparing the rates of absorption in the different classes of cones the visual system is able to discriminate wavelength. These three wavelengths also provide the basis for trichromacy as one approach towards describing human color vision.

Another interesting phenomenon is that these three cone types do not appear with equal frequency, there are 40 red to 20 green to 1 blue and in the foveal area there are almost no blue cones. This implies that the probability of absorbance not only

varies with wavelength but also on the relative distribution of cone types.

As mentioned earlier, the electro-chemical reaction in the photopigments in the rods and cones results in a change in electrical potential of approximately $2\mu V$ across the cell membrane in the outer segment of the rod or cone. This signal arrives at the base of the rod or cone $2ms$ after the light is absorbed by the retina. By the principle of univariance, the cells electrical polarization and hence, its electrical output increases with the rate at which photons are absorbed. It is interesting to note that there is no information about the spectral content of the incoming radiation this information has to be regenerated, and is thought to be a post receptoral activity.

2.1.2 Encoding

This process starts with electromagnetic waves from the environment passing through the cornea, the anterior chamber, the lens, the vitreous humour, the macular coating, the ganglion cells, the amacrine cells, the bipolar cells, the horizontal cells, then finally to the rods and cones that are capable of absorbing this energy. These components all interact with the incoming energy waves in their own way.

After initial stimulation of the sensing elements in the retina, the resulting signals are transmitted to the layer of nerve cells in the retina here some pre-processing, signal conditioning and compression is done on the signals before they are transferred along the optic nerve to the visual cortex. This is inferred from the fact that their are approximately 150 million rod and cone receptors while only about 1 million ganglion cells in the optic nerve. A simplified explanation of this process as currently understood is now given. After leaving the rods and cones the signals travel to a layer of cells just in front of the retina which consists of three types of cells these

are horizontal, amacrine and bipolar cells. These cells are responsible for the initial coding of the visual stimulus before it is sent to the brain and can be seen in Figure 2.2.

The horizontal cells behave differently than other neurons in the body. The typical output of a neuron is somewhat digital in nature (even though it is an analog device). as there is typically a brief electrical activity (called a spike discharge) and the information about stimulus intensity is conveyed by the frequency of the discharges. The horizontal cells, on the other hand, typically do not behave in this manner as their outputs are not spike discharges but rather graded potential changes consisting of an increase in its normal resting potential (hyperpolarization) or a decrease in its normal resting potential (depolarization). These changes in potential are proportional to the stimulus intensity. These potentials have been demonstrated in primates and are called S-potentials. There are two types of horizontal cells, L-type and C-type. The L-type cells are thought to code information about luminosity while the C-type codes wavelength information. The L-type response originates from all three types of cone receptors while the C-type responses results from combination of R and G or R,G and B.

2.1.3 Transfer

In about $50ms$ after receiving the first photon stimulant this information is passed from the retina along the optic nerve to the base of the brain. This is where all the truly wondrous processes begin to happen. First the stimulus leaves the eye and travels along the optic nerve to the brain. The signals from the left and right side of the retina go to the left and right side of the brain respectively with some overlap in

stimulus from the macular region. The images in the two eyes are slightly different and results in a stereoscopic effect from which we get one cue for depth perception. Some crossing over of the nerves are necessary to accomplish this and occurs in the optic chiasma. These fibers terminate in the lateral geniculate nucleus (LGN) a short distance away from the optic chiasma and forms synapses with the other fibers leading to the other parts of the brain concerned with vision.

As mentioned earlier, there are also the amacrine and bipolar cells present just beyond the retina before the ganglion layer; their organization and structure is shown in Figure 2.2. The ganglion cells could share outputs from several cones but luminance and color information are thought to remain separate. This process is known as convergence where there is some coding, integration and feature extraction is done before the information is passed to the other parts of the brain for processing. This is inferred from the fact that there is approximately 150 million sensors in the retina but only about 1 million ganglion cells. This does not indicate receptor redundancy but appears to be an integral part of the coding process which allows the visual system to adequately code information about luminance, form, movement and color.

Convergence on a ganglion cell occurs over well defined areas of the retina. These are called receptive fields and have characteristic spatial distributions. There are also opponent behavior in these receptive fields where the presence of light in one area will inhibit the cell response in a surrounding area. This is the reason one perceives a heightened contrast at a luminance boundary. At higher levels of illumination the effective field size is reduced therefore the number of functional receptive fields per unit area is increased allowing for the discrimination of finer detail.

Luminance is not the only information coded as an opponent signal as color is

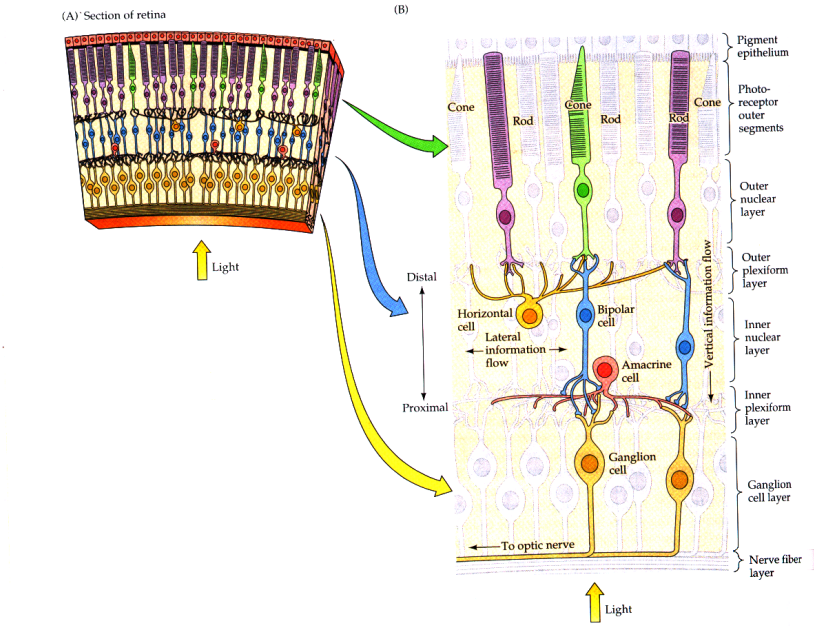

Figure 2.2: Detail of retina showing horizontal, bipolar and amacrine cells [25]

also coded in a similar way by ganglion cells.

2.1.4 Processing in the Brain

The activities related to image interpretation and decision making take place in the brain. A simplified description will now be given derived from [25]. After the rods and cones are stimulated the output signals are then processed and sent as electrochemical signals through the ganglion cells to the lateral geniculate nucleus (LGN) and finally to the visual cortex where analysis of these signals take place. There are two paths that have been identified for this signal transfer from the ganglion cells, these are, the magnocellular and parvocellular streams from what are termed ganglion M and P cells respectively. Experiments indicate that the magnocellular streams conduct information that is critical for the analysis of motion while the parvocellular streams handle information critical for the analysis of shape, size, and color. It has also been observed that the response of the parvocellular cells are slower than the magnocellular cells [26]. It would thus appear that the problem of current interest would be addressed through the processing of parvocellular signals.

Anatomical and electrophysical studies in the monkey have lead to the definition of several functional areas in the brain. The signals leave the retina through the optic nerve to the LGN (Lateral Geniculate Nucleus) from which they are parcelled out to other areas of the brain defined as V1, V2, V3, V4, and MT or the middle temporal area. MT for example,contains neurons that respond selectively to the motion of edges. V4 on the other hand responds to color without regard to motion.

2.2 Theories of Color Vision

The three theories that currently govern models of vision are the trichromatic theory, the opponent theory and the retinex theory. No single theory currently describes all of the known properties of the human visual system and so a composite theory is usually assumed.

2.2.1 Trichromacy, Opponency and Retinex

The theoretical basis of trichromacy could be said to have been started with Newton who first asserted that light itself is not colored but this was rather an interpretation that the brain placed on the distribution of the incoming radiation. Thomas Young was the first to propose that there were three types of sensors in the eye. His thesis was based on the fact that there were seven primary hues (VIGBYOR or violet, indigo, green, blue, yellow, orange and red) and that it would take at least three sensors to represent this combination of hues. In 1855 Maxwell was able to show that the spectral colors could all be obtained by mixing 3 primary colors. Helmholtz was then able to support this experimental data with a physiologically based hypothesis of three channels with different but overlapping spectral sensitivities.

The opponent theory was proposed to explain the phenomenon of opponent colors that is observed in the HVS. Specifically we never seem to observe colors that are combinations of red and green or yellow and blue. Cells have been found in the eye that respond to these opponent colors. They are thought to allow for efficient coding of color information as there is currently significant overlap between the L and M sensor responses and the opponent responses are less correlated. Other kinds of

behaviors are also observed in the human cells that respond to light. One of the more significant is that of wavelength opponency in which the cells respond based on the actual spectral distribution of the incoming light energy.

Wavelength opponency is exhibited by behavior of the horizontal cells in the retina. Several researchers have documented Hering's theory in which he identified a red/green and a blue/yellow opponent behavior in the human visual system. Hering also described a light to dark opponent system but a pathway for this signal has not yet been found. This theory is observed from the fact that one is not able to see colors that are a mixture of red and green or colors that are mixtures of blue and yellow, these colors are thus said to be opponent. It should be remembered however, that the names red, yellow etc. are our descriptions for light reaching our eyes with particular wavelength distributions. These responses are more accurately described in terms of the response of sensors in the eye that are sensitive to long, medium and short (LMS) wavelength radiation.

It is thought that this opponent coding mechanism is an approach to code the signals travelling from the retina to the brain that helps to reduce the correlation between the L and M sensors because of their significant spectral overlap. This then would allow for more efficient coding in the brain. The hue cancellation experiment produced curves that showed what these spectrally opponent responses looked like. These are shown in Figure 2.3 denoted as $F_{rg}(\lambda)$ and $F_{by}(\lambda)$, to represent the red-green and the blue-yellow opponent responses respectively.

Everyone has experienced the phenomena of color constancy. As we move from the interior of a building to the outside the spectral distribution of the light impinging on our retinas changes. This should result in a change in the color of objects as we

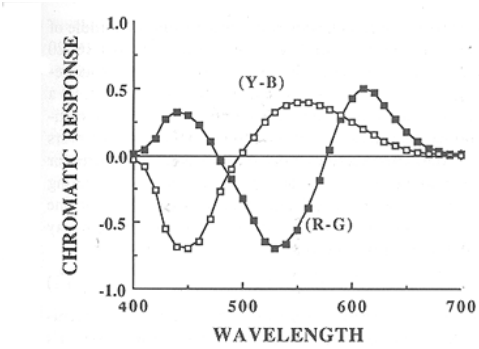

Figure 2.3: Wavelength opponent response of ganglion cells in the retina [27]

move from inside to outside. This is not the case under most circumstances and the ability of the HVS to accommodate these changes is called color constancy. Land [24] developed a theory called the retinex theory to explain this phenomenon. In it he proposes that this information is coded in terms of the relative lightnesses of the objects in the scene as determined from the sensor responses. This in effect allows the visual system to extract information about the reflectances of the objects in the scene which is the only property that is not changing.

2.2.2 Relating the theories

Currently there is no equivalent to the unified theory for explaining visual phenomena. As a result models have been developed to explain various observations. The trichromatic theory while doing a satisfactory job of explaining color matching does not explain the existence of opponent colors thus the opponent theory. Neither of these two theories do a satisfactory job of explaining color constancy and appearance and thus the rise of Land's retinex theory. It would thus appear that the truth lies

in some combination of the above theories.

2.3 Models of Visual Information Processing

A receptive field, is a visual area within which light influences a neuron's response. This is a ubiqitous apparatus in the human visual system and occurs along all parts of the visual chain from the cells in the retina to higher level regions of the brain. It appears that these receptive fields are particularly suited for processing and representing contrasts in a behavior that is locally linear (i.e. they display the properties of a linear system around particular levels of illumination). It therefore appears that much of the representation in the brain relies on contrasts and are therefore significant features to be utilized by any image processing system.

Receptive fields typically have a center surround geometry as shown in Figure 2.4. There are three types defined in [24], an on center off surround (Type I) and off center on surround (Type II) and a Type III which responds like the photopic luminous response of the eye. In addition there are thought to be double opponent cells in which the center and surround responses are driven by combinations of the LMS sensors. The receptive fields identified in Figure 2.4 in combination in, effect behave like band pass filters of varying bandwidths and end up comparing image responses on different scales. The outputs of these receptive fields could also be coding lightness contrasts thought to be a key requirement for color constancy.

It is thought at this time that these fairly simple recognition tasks are conducted at the lower levels of the brain and that there could even be feedback, learning and memory that affects the process so that the decisions could possibly be made at the

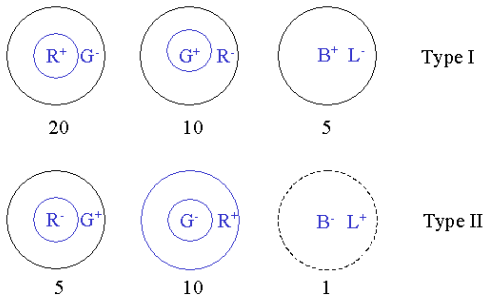

Figure 2.4: Ganglion receptive field showing spatial opponency with a center surround relationship identified in the eye [24]

ganglion or the LGN. Much of the coding of images relies on contrasts and edges which can be obtained from the outputs of receptive fields. Marr [7] defined the purpose of vision as that of representing shape. In order to get to that description, however, we need to describe edges and surfaces. He then goes on to describe filters that support this process. A filter that is believed to be significant in this process is the $\nabla^2 G$ operator called the Laplacian of the Gaussian where ∇^2 is the laplacian operator and G is a gaussian kernel as defined in Equation (2.1) and Equation (2.2).

$$\nabla^2 \equiv \partial^2/\partial x^2 + \partial^2/\partial y^2 \tag{2.1}$$

$$G(x, y) = e^{-(\frac{x^2-y^2}{2\pi\sigma^2})} \tag{2.2}$$

These can be used to detect intensity changes at many different scales based on the choice of σ in Equation (2.2). Marr also went on to show that under certain conditions the $\nabla^2 G$ operator could be approximated as a difference of gaussian (DOG) functions which would describe the neuron receptive fields.

2.4 Summary

The sensing of color by humans could be characterized in three stages. Stage 1 would be sensing as occurs in the eye. Stage 2 would be coding and transfer to the brain and Stage 3 decoding and interpretation of the data. The major activity that forms our sensing of color is done in the brain. Again, if we knew the forms of these representations could they be of use in the design of artificial systems. Interpretation of what

we see is probably the darkest art and probably relies heavily on apriori knowledge and experience. In the end the parameters needed are coded and sent to the visual cortex for further processing. These are all coded in the systems described earlier. The mechanisms for extracting and utilizing this information then seems to reside in the visual cortex. The science of color must be regarded as a mental science James Clerk Maxwell 1872 [28]. We are interested in the use of these representations especially at the early stages of coding and their potential use to guide the development of algorithms.

Chapter 3

Biological Operation Based Vision (BOBV)

In this approach we will describe three stages that correspond to the three stages of processing identified in the human visual system as described in the previous chapter. They will be called levels, instead of stages, to differentiate the artificial and approximate process to be developed here from the natural process in the human visual system.

The motivation stems from the following facts:

1. Humans are very good at vision.
2. Many problems of interest are done very well by humans.
3. We can learn principles that would be applicable to other signal processing domains.

Specifically, we obtain through user descriptions, biological models, and physical measurements these salient features which are then used to streamline the formulation of algorithms; thereby allowing for efficient convergence to effective algorithms.

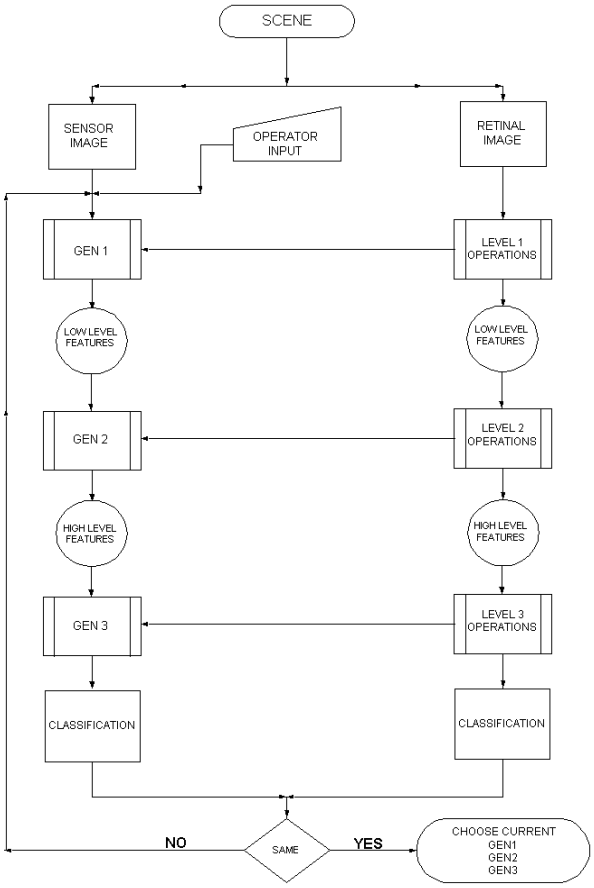

Figure 3.1: Formulation Process for BOBV Algorithms

3.1 Overview

What is envisioned here is an operator driven algorithm development process whereby a combination of operator inputs, and psychophysical models are used to drive the development of algorithms. The overall process is illustrated in Figure 3.1, where GEN1, GEN2 and, GEN3 represent the procedures corresponding to the operations at levels(1,2,3) in the human visual system (HVS); these are described in more detail below.

This technique should produce algorithms that would be implemented at the acquisition/preprocessing level, low level vision operations, and higher level vision operations, denoted as levels 1, 2 and 3 in Figure 3.1. The methodology will allow the user to directly influence the selection of the appropriate algorithms. It is envisioned that the system would consist of four major functions as follows:

1. Level 1 Operations (GEN1).
2. Level 2 Operations (GEN2).
3. Level 3 Operations (GEN3).
4. Evaluate and Modify above as necessary(Feedback).

We assign to LEVEL 1, operations that would occur in the ganglion cell layer in conjunction with operations in the LGN (Lateral Geniculate Nucleus) as described in Section 2.1.4. LEVEL 2 activity, would correspond to the activity in V1 while LEVEL 3 would cover operations at the higher levels of the brain corresponding to what is currently defined in areas V4 to MT. Receptive fields have been identified at all these levels with differing functionality [29]; they respond in general to the spatio-temporal, chromatic as well as the binocular elements of the signal. We seek, in this approach, to exploit the representations that result from these operations as features for classification.

The overall process would function as follows: the paths identified in Figure 3.1 show two parallel processes, a machine process on the left, and the human process on the right. They both begin with a representation of the scene of interest on an image sensor and the retina respectively. In the human process, the next stage is the LEVEL 1 operations which leads to the extraction of low-level features. This includes the features used for coding of the image data before transfer to other parts

of the brain. Knowledge of these operations will be used to drive the development of methods in the box labeled GEN1 on the machine side.

As we continue down the HVS path these low-level features are passed to the parts of the brain defined as conducting LEVEL 2 operations as would be conducted for example in V1. At this stage, higher level features are generated, these would include things such as color, shape, motion, or texture. Like the previous stage, we would use knowledge of these operations to drive activities in GEN2.

The next stage would be purposive meaning that the operations would be driven by the goals of the process; so for example, the operations needed for tracking a moving baseball are different than those for discerning a change in color. These ideas, however, would drive the operations developed in GEN3.

The situation when dealing with the typical inspection problem is less complicated than the general vision problem as usually there are only a few items of interest at any one time. The box labelled operator input is used to identify the elements that are of interest in the scene. The other salient feature is that of feedback where we iterate through the combinations of features and operations to get a sequence of operations to meet the needs of the application.

We have outlined a general approach that can be enhanced as knowledge of the operations in the human visual system improves.

The general scenario for imaging is shown in Figure 3.2 where you have a light source, and object(s) to be imaged along with a system for acquiring the image which could be a camera or the human eye. A model for the development of imaging algorithms under these general conditions will now be presented.

First we describe the assumptions which follow, these are:

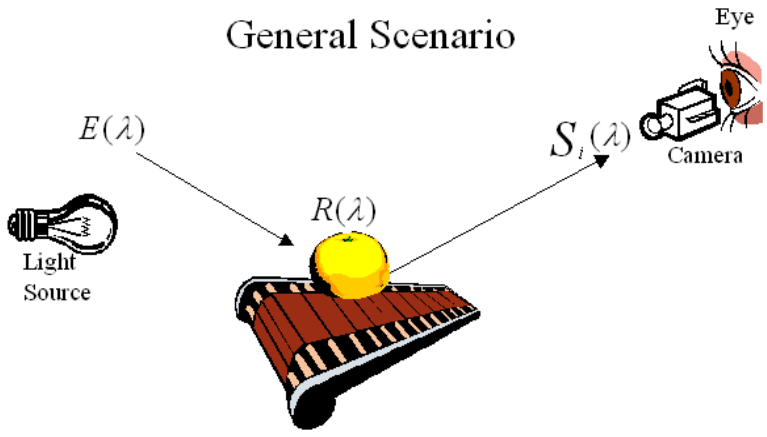

Figure 3.2: Common configuration for industrial imaging

- Photopic vision (vision driven by the cones or bright light vision)
- Monocular vision (stereo effects are not significant)
- No motion cues (detection of motion in not necessary)
- Depth cues from apriori knowledge and shading (can tell defects by looking at an image on a monitor)
- No specular reflection (body reflection to represent colors as opposed to the wavelength distribution of the light)
- Low level image processing (operations are conducted at high rates)

The assumption of low level image processing stems from the fact that for most of these applications the decision making is done at very high rates which would lead us to believe that the processing is being done very early in the visual stream. Much of this processing seems to rely on our ability to detect contrasts between the

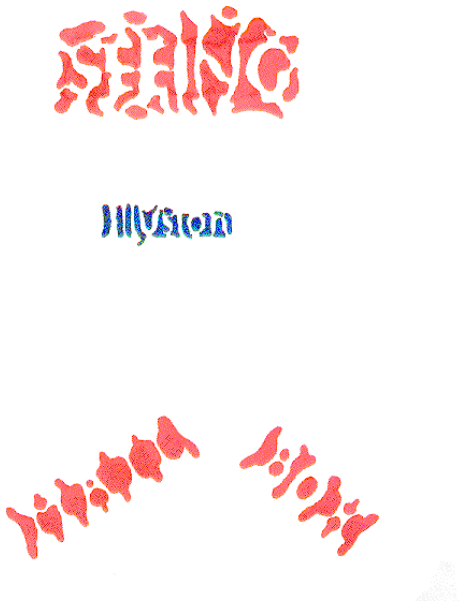

Figure 3.3: Example image to differentiate seeing from perception

normal and abnormal. This appears to be relatively easy for most humans making decisions on the production line, the main problems stems from their inability to maintain a consistent level of concentration for extended periods of time for example over an 8 hour shift. An example of the kinds of operations to be considered can be gleaned from Figure 3.3 we would be more concerned about detecting the blobs and their presence as opposed to being able to decipher the words in the blobs. This should thus be considered more of a Machine Vision as opposed to a Computer Vision problem [1] driven by more practical and time critical considerations.

3.2 Contrast and Receptive Fields

Contrast appears to be an important element of human visual processing. We will now describe the phenomenon of contrast, as well as receptive fields, which are designed to operate on and enhance contrast.

3.2.1 Importance of Contrast

We will define contrast in this context as a deviation of a signal from some average or background value. This is different, for example, from another common definition in which it is defined as the extent of variation in an image or the dynamic range of the image (this is what is usually adjusted with the contrast adjustment on your TV or computer monitor). In general, our brains seems to try and increase contrast between targets and the background. It appears that the HVS continually strives to enhance localized contrasts.

One way to view this is as a difference operation. This can be accomplished in several ways and might be problem specific so that usable contrasts are learned for the specific problem under consideration. For example, in one case the contrast might be one of texture, while in another it is an edge contrast, while in another it could be a color contrast. These operations are also considered to be LEVEL 2. The general effects of these operations is to enhance contrast in the image. Contrast appears to be one of the major mechanisms used by the human visual system to conduct fast robust detection and identification. This is inferred from the fact that contrasts stretch (or expand) the large dynamic range we can see, which spans about six orders of magnitude in terms of incident energy, from a dim evening to a bright

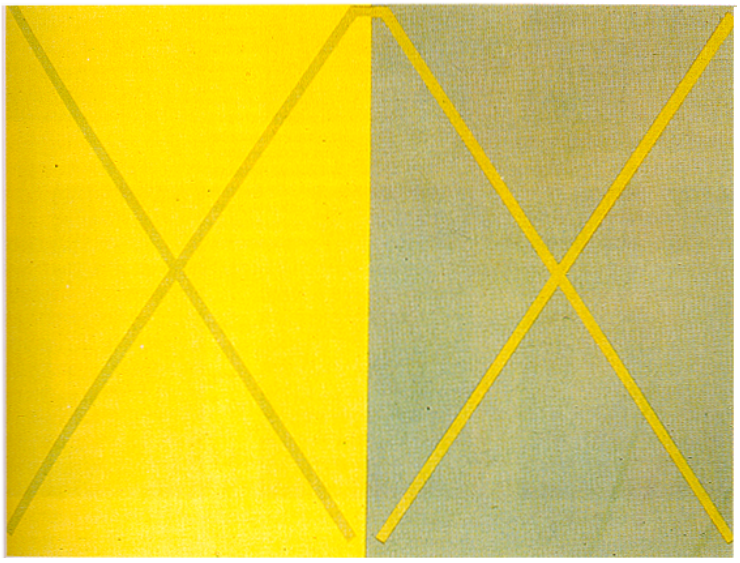

Figure 3.4: Effect of contrast on appearance [26]

sunny day. The individual neuron response can only handle about two to three orders of magnitude change. Contrasts in natural scenes, vary on the order of two to three orders of magnitude more directly matching the neuronal responses. It is also felt that the information in images is the contrast [30].

Additionally, contrasts are closely related to the properties of surfaces and is usually the information of interest. Contrast, however, is not directly related to appearance but is more concerned with our ability to locate areas of interest in a scene. Much of this capability is governed by the scene itself as illustrated in the Figure 3.4 where we are able to discern the quite easily the differences brought about by the cross even though they appear different (note the x in the image on both sides have the same radiometric properties).

Let us look initially at simple scene as shown in Figure 3.5(a). For this scene

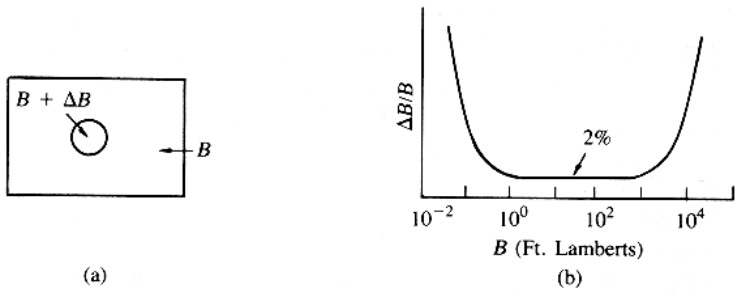

Figure 3.5: Contrast in simple scenes [31]

we have an object of uniform color on a uniform background. Figure 3.5(b) shows the detectable contrasts where ratios above the line are detectable by the HVS and ratios below are not. This threshold for detectable contrasts is called the Weber ratio as described by Gonzalez[31]. This implies that for the simple scene luminance contrasts of 2% or greater are detectable within certain overall illumination limits. Outside of these limits—which fall outside the range for photopic vision— higher ratios are required

The situation changes significantly in more complicated scenes, where the background illumination has a substantial effect on the detection thresholds as shown in Figure 3.6. It can be seen that the behavior falls within the limits of the detection thresholds for the simple scene shown in Figure 3.5. These definitions concern luminance thresholds and do not tell us about color. Color is of importance, however, as in most natural scenes the ability to discern differences in the scene depends on color.

Using the definition of contrast as defined in Equation (3.1) and shown in Figure 3.5

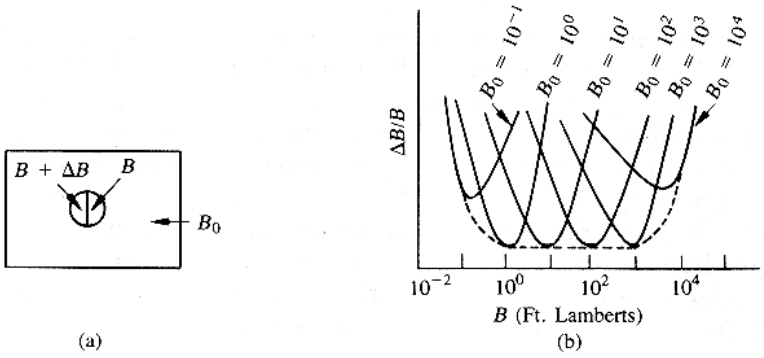

Figure 3.6: Contrasts in complicated scenes

$$C(x,y) = \frac{\Delta B(x,y)}{B_{ref}} \tag{3.1}$$

where C is defined as the contrast, ΔB is a difference with respect to a reference, and B_{ref} is a reference value. The choice of the reference value will be discussed further below but is based on the overall characteristics of the scene. Using Equation (3.1) it is possible to identify in an image pixels where the contrast threshold is exceeded as in Equation (3.2) these pixels could then be used to identify areas of interest in the scene for further analysis.

$$C(x,y) \geq C_{\min} \tag{3.2}$$

For a color image we would write Equation (3.1) in a more general form as shown in Equation (3.3) where i defines the contrast in each of the sensing planes. As a

result there will now be not only a magnitude but also a direction for this contrast. These additional degrees of freedom will assist in the process of classification.

$$C_i(x,y) = \frac{B_i(x,y) - B_{i\,ref}}{B_{i\,ref}} \tag{3.3}$$

Additionally we would also define a more general contrast threshold as in Equation (3.4) to be used for thresholding as the value of 2% given in [31] is for a luminance contrast. The weights w_i given in [32] ($w_1 = 0.299;\ w_2 = 0.587;\ w_3 = 0.114$) for example, gives a better correspondence with the perceived brightness of color, and would allow for the computation of an equivalent luminance contrast C_{tot}. This knowledge could be used for example to filter noise in images that approximate simple scenes; as pixels with these values would not be observed by the graders in many applications.

$$C_{tot} = \sum_{i=1}^{3} w_i C_i(x,y) \tag{3.4}$$

The visual effect of contrasts computed with Equation (3.4) for model generated sample images are shown in Figure 3.7. The images consists of an elliptical region in the center on a uniform background. Figure 3.7(a) is above threshold with $C_{tot} = 0.032$, Figure 3.7(b) is right at threshold with $C_{tot} = 0.025$, while Figure 3.7(c) is below threshold with $C_{tot} = 0.011$. This shows reasonable agreement with observations.

3.2.2 Encoding/Processing Contrast with Receptive Fields

It is believed that the receptive fields in the ganglion layer of the retina is a significant mechanism for encoding luminance contrast. As mentioned earlier however receptive

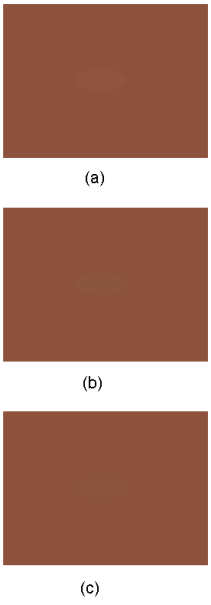

Figure 3.7: Images with varing color contrasts (a) 0.032, (b) 0.025, (c) 0.011

fields occur at many places along the signal paths to the brain and are sensitive to spatial, temporal, chromatic as well as binocular (things that are seen in either eye). The receptive field model, assumes that the neural response is due to two separate mechanisms called the center and the surround with the center being of a smaller spatial extent than the surround. A general equation to describe this behavior is given in Equation (3.5) which describes a response R_i, where α_c and α_s denotes the weights for the center and surround respectively, x, y indicate spatial position; t time, and z the disparity for binocular images. I_c and I_s identify the inputs for center and surround processing with $*$ being the convolution operator.

$$R_i = \alpha_c Zc(x, y, z, t) * I_{ci}(x, y) - \alpha_s Zs(x, y, z, t) * I_{si}(x, y) \tag{3.5}$$

It has been shown that the responses for some cells are separable [33] so that we can write Equation (3.5) as shown in Equation (3.6).

$$R_i = \alpha_c Z_1 c(t) Z_2 c(x, y, z) * I_{ci}(x, y) - \alpha_s Z_2 s(t) Z_2 s(x, y, z) * I_{si}(x, y) \tag{3.6}$$

Looking only at monocular responses we would then obtain the response output as in Equation (3.7).

$$R_i = \alpha_c Z_1 c(t) Z_2 c(x, y) * I_{ci}(x, y) - \alpha_s Z_2 s(t) Z_2 s(x, y) * I_{si}(x, y) \tag{3.7}$$

Next, assuming no temporal responses we would obtain the response as shown in Equation (3.8) which describes the responses of the simplest receptive fields.

$$R_i = \alpha_c Z_2 c(x,y) * I_{ci}(x,y) - \alpha_s Z_2 s(x,y) * I_{si}(x,y) \tag{3.8}$$

Using Equation (3.8) and choosing Z_2c and Z_2s to be Gaussians we obtain the Difference of Gaussian (DOG) model. The use of the DOG model for describing the behavior of the receptive field has been shown to be useful in describing the behavior of ganglion cells [29] [13]. The curves describing the spatial sensitivities of these two areas are assumed to be Gaussians as these have been shown to be representative of the human responses and additionally reduce the effects of ringing and filter ripple when used in the convolution operations.

The composition of the signals that make up the center and surround input is still a matter of debate but several scenarios from various anatomical studies have been presented. Some of these from [7] are shown in Figure 3.8. A summary of many of the proposed combinations are presented in Table 3.1 where the center input is defined as I_{ci} and the surround input as I_{si}. The past research propose many forms for the I_{ci} and I_{si} that are typically combinations of the long, medium, and short (LMS) wavelength sensors in the eye. These are also typically referred to as R,G and B sensors.

We will define center and surround responses based on a combination of the trichromatic and opponent theories of vision. The responses based on the trichromatic responses will be called $Class\ I$ and those due to the opponent responses will be termed $Class\ II$. These will be described more completely in Chapter 4. The main point to note are the simplifying assumptions made to obtain the DOG model, indicating that we are utilizing a limited spectrum of the behavior of these receptive

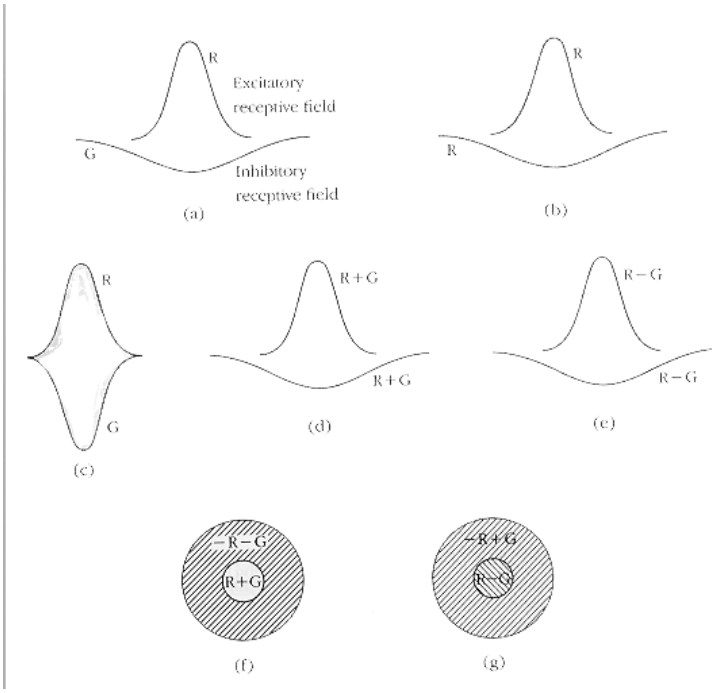

Figure 3.8: Receptive fields as described by Marr [7]

Table 3.1: Classes of Responses

i	I_{ci}	I_{si}	Source
1	R	R	[7]
2	G	G	[7]
3	B	B	Postulated
4	R	G	[7]
5	R	R	[7]
6	R	G	[7]
7	R	G+R	[34] [26]
8	G	G+R	[34] [26]
9	B	G+R	[34] [26]
10	-(R+G)	(R+G)	[7]

field responses and the information they encode.

We have identified here one of the significant characteristics of the human visual system for exploitation in the development of machine vision algorithms. Many more general models of the functions of the human visual system has been developed driven by other motivations such as target recognition [10]. In a more general sense, another approach to this problem would be to look at simplifying some of these more complicated models to meet machine vision needs.

3.3 Problems of Interest

Computer Vision algorithms and especially those derived from human visual models have not typically been applied to machine vision problems. There are many reasons for this:

1. Typically in many other domains of computer vision applications, there is also usually a human in the loop for example in applications such as analyzing satellite or medical images. A radiologist looking at x-rays, for example can, in most cases, take the required time to make a determination on what's in an image. During operation of these systems the human assists in making the final determinations.

2. Usually, the people using these systems are also fairly experienced in the use of digital imaging and the common image processing techniques and are able to choose the sequence of operations that are germane to the problem under consideration. This is not usually the case in machine vision problems where

one usually has people that are expert at conducting a particular operation in a manufacturing environment but are not usually familiar with image processing.

3. Most general image processing algorithms (including those derived from human models) are typically applied in operations that are usually not time critical. These algorithms execute on the order of minutes as opposed to seconds or fractions of a second for most machine vision applications in food processing.

4. Most machine vision solutions today are also still monochrome and those that use color typically treat the output as a group of gray scale images. This does not normally meet the requirements of most food processing applications.

For the applications of interest here, techniques that are able to derive useful solutions with a prescribed approach would be of much help to the machine vision system designer. In particular, for machine vision applications in food processing the human involvement would mostly be in training and not in normal operation. We will present a unified approach for processing color images based on the functions of the human visual system as currently understood. This approach enables us to develop routines that conduct fast color segmentation and classification to meet the above described needs.

Problems with the characteristics typical for natural products are shown in Figure 3.9. In the first two examples shown in Figure 3.9(a) through Figure 3.9(d) we are interested in detecting the fan bone. In the latter two examples, shown in Figure 3.9(e) and Figure 3.9(h), we are required to categorize the overall surface condition of the grapefruits; this could be succinctly described as determining the amount of surface area that is not yellow-green. Additionally, for this latter application it might

also be required that some absolute color tolerance be observed. In general, for both problems, the overall goal could be described as transforming several thousand bits of information to just a few that indicate the status of the part.

The fan bone problem is driven by a need to fuse information from two modalities, visible and X-ray images. This is necessary because for the X-ray system the energies needed to detect the hard sub-surface bones are not able to easily detect the softer surface bones such as the fan bones. A combination of both modalities, however, could serve to enhance the overall system accuracy. We also show in Figure 3.9 the original color images on the left with the gray scale versions of the same images on the right. First, it should be observed from the gray scale version of the images, that the defects of interest cannot be detected as they would be equivalent to shadows and shading. For example, with the fan bone images it would be difficult to separate the bone from the background. This then makes it necessary to process in color space increasing by a factor of three the data that must be handled. In addition, there are subtle changes in color that have to be detected so that we do not confuse the bruised region with the fan bone region as shown for example in Figure 3.9(c). People currently conduct the fan bone inspection tasks, but they are error prone, as it is difficult to maintain their attention for extended periods; especially at the current line rates of about 35 parts per minute. A system that could conduct the fan bone screening in real time could significantly improve the efficacy of the overall inspection operation.

Similarly, for the grapefruit sorting, it is necessary to identify problems due to surface defects and discolorations. This problem is also currently done by people but at rates of 600 parts per minute, much higher than those for the fan bone problem.

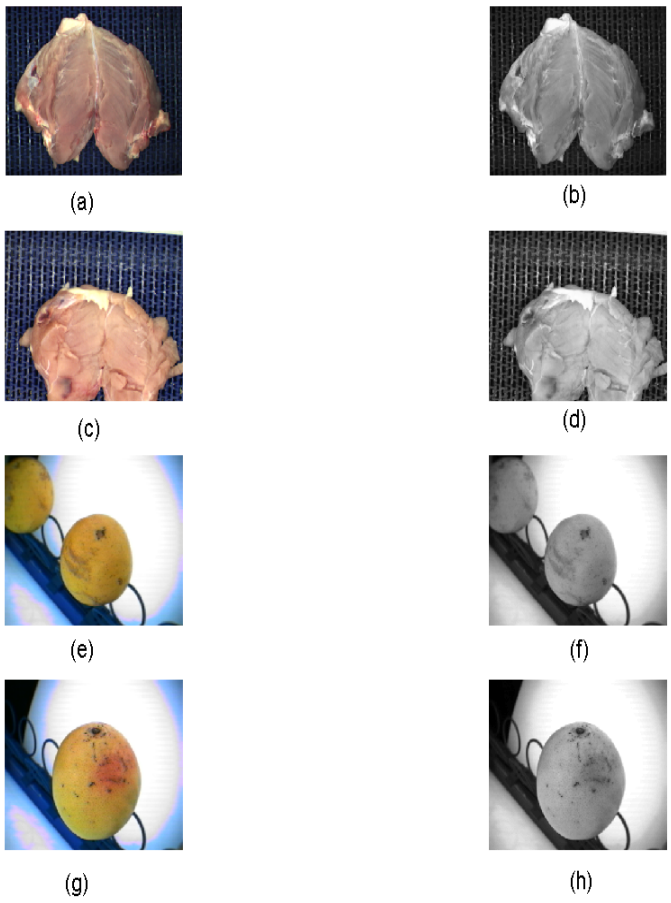

Figure 3.9: Characteristic problems when sorting natural products

It is also observed that the gray scale images do not accurately reflect the surface conditions of the fruit as the blush seen in Figure 3.9(g) would easily be confused with shading and shadow in Figure 3.9(h). Additionally, the guidelines for screening (as with other natural products) are somewhat subjective in nature so that an automated solution would provide more consistency in the determinations.

3.4 Summary

In this chapter, the essential elements of the human visual system have been identified as they relate to the problems under consideration. Specifically we have identified contrast as a significant mechanism in the process and proposed a technique for calculating color contrasts. We have also outlined in this chapter the general approach to be used and framework within which the operations to guide the development of machine vision algorithms will be conducted. Additionally, we have outlined the structure and function of the general receptive field and the assumptions to derive the DOG model. Also, we have presented two problems of interest as motivation for the development of the approach to be described. The next step will now be to look into more detail at the mechanics involved in implementing the approach using the DOG model and to determine its behavior under different conditions.

Chapter 4

Approach to the Extraction of Features Using Contrast

In Chapter 3 we proposed approaches for image feature extraction using bio-physiological descriptions of the human visual system operations as the basis for obtaining features. In this chapter we will perform an analysis of the approach to identify the significant characteristics of the approach and to make a determination of expected performance. Our basic assumption is that contrast is a significant element of the process and that an approach using this method leads to the generation of more robust features. In addition a somewhat general approach to the development of vision algorithms is proposed.

4.1 Procedure Overview

The human visual system is a remarkable engine for the processing of visual information. Using the knowledge about the human visual system summarized in Chapter 3, we now look at approaches to use this knowledge for the generation of algorithms. The specific problems of interest are inspection problems that are currently done most effectively by humans. For these problems it is common practice to train inspectors through the use of examples; that is, they will usually be shown both good and

defective product and then through experience learn what is classified as defective product. We will follow a similar approach by identifying the significant features that would be computed by the human visual system (HVS) and how they could be used in discrimination. Factors or features that appear to be of importance in doing this evaluation include:

- Edge and boundary detection (transition areas in images identify material or other changes)
- Detection of colors (most problems of interest cannot be solved satisfactorily with monochrome images)
- Ability to tolerate noise (some defects can be similar in appearance to normal areas)
- Some invariance to spectral shifts (slight shifts in intensity or wavelength distribution are accommodated)

The signals and processes we can infer that are significant in obtaining these features include:

- Long, Medium and Short (LMS) responses
- Spectral opponent responses
- Receptive fields

The assumptions for the analysis as identified in Chapter 3 are as follows:

- Photopic vision: most inspection tasks are conducted at levels of light that would put us in the photopic (vision mediated by cones) regime.
- Monocular vision: while more difficult it is possible to do these tasks with one eye.
- No motion cues: object does not have to be in motion to observe the defects.
- Depth cues from apriori knowledge and shading: based on monocular assumption above
- No specular reflection (body reflection): we are not able to see defects on specularly reflecting parts of the scene.
- Low level image processing: the rates at which the typical tasks are conducted indicate very fast processing as characterized in the previous chapter.

4.2 Mathematical Formulation

In defining the response for a receptive field the general equation could be written as shown in Equation (4.1)

$$R_i(m,n) = \alpha_c \beta_{ci}(m,n) * I_{ci}(m,n) - \alpha_s \beta_{si}(m,n) * I_{si}(m,n) \tag{4.1}$$

where β is a convolution filter; $*$ denotes the convolution operator; I denotes an input image; and the subscripts "c" and "s" denote center and surround respectively. The main task will then be in determining the combination of features that optimizes the

process of discrimination. This implies determining the combination of outputs ($R_i's$) and their relevant parameters ($\alpha's$ and $\sigma's$) that maximizes the contrast between a defect to be identified and the surround. We now look at the use of some of these models for the generation of features that could be used for classifying image regions for the purpose of identifying defects, specifically through the use of the receptive field concept.

The problem then is to identify possible responses R_i that are significant for a given problem, represent the problem in this space and do cluster analysis for segmentation or object identification. This then becomes an optimization problem:

$$Max(|\mathbf{R}_b - \mathbf{R}_t|) \tag{4.2}$$

where $\mathbf{R}(m,n)$= $\mathbf{f}(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$ is formulated by rewriting Equation (4.1) in a more compact form, while b and t represent the background and a target or object of interest respectively. We then need to choose parameters $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ to satisfy Equation (4.2) if the problem is one of segmentation.

If the problem is one of edge location then Equation (4.3) where ($\mathbf{R}_b \longrightarrow \mathbf{R}_o$) indicates a transition from a background area to a region of interest and ξ an operator to locate an edge in this transition. Our task then would need to choose parameters to satisfy Equation (4.3). In some applications, it could also be desirable to utilize a combination of these two features (regions and edges).

$$\xi(\mathbf{R}_B \longrightarrow \mathbf{R}_O) = \xi_{TRUE} \tag{4.3}$$

The $\alpha's$ would serve to describe the relative weights of the center surround response while the $\sigma's$ are the parameters for the Gaussian filters. The algorithms is then developed based on the principle of BOBV and the detection of contrasts.

The idea here is to utilize the Human Visual System (HVS) as a model for developing machine vision algorithms to solve problems that are typically addressed by humans in an acceptable manner. We now look at the behavior of some of the biological mechanisms that are in place in the HVS and their influence on these operations.

The problems to be considered are general in nature for people conducting sorting or other quality control functions. We identify three possible tasks that are of potential interest for machine vision, and look at developing a set of features/characteristics from the image that allows for robust segmentation and classification. These tasks are: (1) Separating a region on an object from the background of that object (this could be termed a detection problem). (2) Identifying a region or objects from other regions or objects (this could be termed classification). (3) Enhancing object geometry or morphological properties (for example size and shape).

As an example consider a 2D (two dimensional) feature set with features $R1$ and $R2$ representing two output responses as shown in Figure 4.1. Feature1, Feature2 and Background represent the clustering of areas of interest in the original scene in the response space as defined by $R1$ and $R2$. For Task 1 we would like to maximize the distances between the cluster centers $D1$ or $D3$, depending on our interest while minimizing the cluster overlap. For other problems, such as might be defined by Task 2 we might have more interest in maximizing $D2$.

We model the response of the receptive fields by using a difference of gaussian (DOG) formulation as presented in the previous chapter by rewriting Equation (4.1)

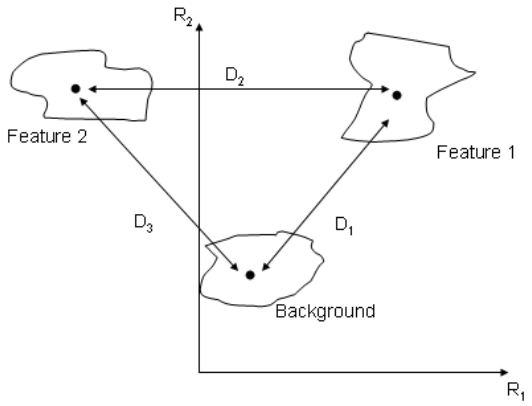

Figure 4.1: 2D depiction of the response space

in terms of the receptive field response with $\beta = \beta(\sigma)$ to get Equation (4.4) where σ describes the spread of the gaussian.

$$R_i(m,n) = \alpha_c \beta_i(\sigma_c) * I_{ci}(m,n) - \alpha_s \beta_i(\sigma_s) * I_{si}(m,n) \tag{4.4}$$

The structure of these filters were shown in Figure 3.8 and Figure 2.4, where the $\beta' s$ are represented as Gaussians with parameters σ_c and σ_s where $\sigma_s \geq \sigma_c$. For the rest of this chapter we conduct the analysis using a $3x3$ kernel. We also set $\alpha_c = \alpha_s = 1$ as they are just scaling factors on the overall responses.

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{m-1,n+1} & \beta_{m,n+1} & \beta_{m+1,n+1} \\ \beta_{m-1,n} & \beta_{m,n} & \beta_{m+1,n} \\ \beta_{m-1,n-1} & \beta_{m,n-1} & \beta_{m+1,n-1} \end{bmatrix} \tag{4.5}$$

Define the filter kernels β to have the form in Equation (4.5) referenced to the center kernel. The filters defined by $\boldsymbol{\beta}$ are derived from the receptive fields. Assume the receptive fields (RFs) are a linear combination of gaussians, we then begin with the Bivariate Normal Density function as shown in Equation (4.6).

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\{-\frac{1}{2}[(\frac{x_1 - \mu_1}{\sigma_1})^2 + (\frac{x_2 - \mu_2}{\sigma_2})^2]\} \tag{4.6}$$

We then center the kernel such that $\mu_1 = \mu_2 = 0$, this produces a filter with zero phase as is desirable to maintain the spatial integrity of the outputs. As an example with a 3x3 kernel and substituting values for kernel positions and assuming gaussians are circularly symmetric ($\sigma_1 = \sigma_2 = \sigma$) we get.

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_2 & \beta_1 & \beta_2 \\ \beta_1 & \beta_0 & \beta_1 \\ \beta_2 & \beta_1 & \beta_2 \end{bmatrix} \tag{4.7}$$

Where

$$\beta_0(\sigma) = \frac{1}{2\pi\sigma^2} \tag{4.8}$$

$$\beta_1(\sigma) = \frac{1}{2\pi\sigma^2} \exp(-\frac{1}{2\sigma^2}) \tag{4.9}$$

$$\beta_2(\sigma) = \frac{1}{2\pi\sigma^2} \exp(-\frac{1}{\sigma^2}) \tag{4.10}$$

The general computation for determining the values that make up the general $\boldsymbol{\beta}$ is shown in Appendix E

Knowing $\boldsymbol{\beta}$ it is now possible to compute output responses for various input images. We will now look at the behavior of these responses.

Define the input image kernel for operations as

$$I(m,n) = \begin{bmatrix} I_{1,1} & I_{1,2} & I_{1,3} \\ I_{2,1} & I_{2,2} & I_{2,3} \\ I_{3,1} & I_{3,2} & I_{3,3} \end{bmatrix} \tag{4.11}$$

Where $I_{1,1} = I(n-1, m+1)$; $I_{2,2} = I(n,m) = center\ pixel$; and $I_{3,3} = I(n+1, m+1)$ *etc*...describe the relative positions from the center pixels.

With

$$\boldsymbol{\beta}_i^T = [\beta_2\ \beta_1\ \beta_2\ \beta_1\ \beta_0\ \beta_1\ \beta_2\ \beta_1\ \beta_2]$$

and

$$Iv_i^T = [I_{1,1}\ I_{1,2}\ I_{1,3}\ I_{2,1}\ I_{2,2}\ I_{2,3}\ I_{3,1}\ I_{3,2}\ I_{3,3}]$$

we can now write the output response as

$$R_i(n,m) = \alpha_{ci}(\beta_{ci}^T \cdot Iv_{ci}(n,m)) - \alpha_{si}(\beta_{si}^T \cdot Iv_{si}(n,m)) \tag{4.12}$$

where $\cdot$ signifies the dot product operation. Equation (4.12) is the form used for computing the output responses.

4.3 Response Function Characteristics

We have to this point described responses based on the receptive fields. At this point we will look in more detail at the general behavior of these responses.

4.3.1 Response Types

We define two basic types of responses for consideration here, $Class\ I$ and $Class\ II$. These have the center surround parameters as shown in Table 4.1 and are chosen

Table 4.1: Definition of Class I and Class II responses

	i	$Center$	$Surround$
$Class\ I$	1	R	R
	2	G	G
	3	B	B
$Class\ II$	4	R	$R-G$
	5	G	$R-G$
	6	B	$R+G-B$

to represent the chromatic ($Class\ I$) and the opponent ($Class\ II$) responses in the human visual system (HVS) based on the color vision theories described in Chapter 2. The basic difference between $Class\ I$ and $Class\ II$ is that for $Class\ I$: $I_{ci} = I_{si}$ and for $Class\ II$: $I_{ci} \neq I_{si}$.

For the $Class\ I$ response, since $I_{ci} = I_{si} = I$, so we can rewrite Equation (4.12) as

$$R_i(n,m) = (\alpha_{ci}\beta_{ci}^T - \alpha_{si}\beta_{si}^T) \cdot Iv_i(n,m) \tag{4.13}$$

4.3.2 Sample Image Analysis

In order to gain some general insights into $Class\ I$ responses we look at a simple example monochrome image consisting of a target in a background as shown in Figure 4.2. This corresponds to a simple scene that is usually used to describe the contrast threshold at which objects are distinguishable from the background. This is commonly known as Weber's Law [31]. We will examine responses for that would represent regions and edges as identified from the problems of interest described in Equation (4.2) and Equation (4.3). First we look at responses for region segmentation.

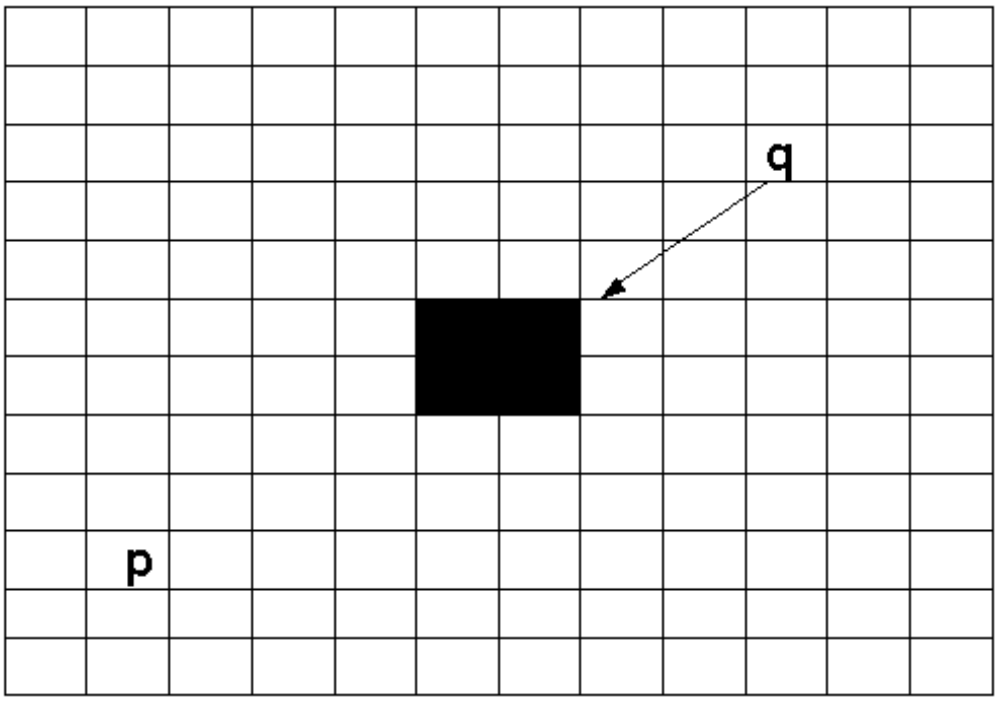

Figure 4.2: Monochrome sample picture input, p background value, q the target area value

4.3.3 Response for a region

We will first describe the overall output response for any area of interest in the image as

$$J = \sum_{md} \sum_{nd} R_i(n, m) \tag{4.14}$$

where md and nd identifies the range of m, n that defines the area of interest or we can write as

$$J = Ctr - Sur \tag{4.15}$$

where Ctr and Str are defined in Equation (4.16) and Equation (4.24)

$$Ctr = \sum_{md} \sum_{nd} \alpha_{ci} \beta_{ci}^{T} \cdot Iv_i(n, m) \tag{4.16}$$

$$Sur = \sum_{md} \sum_{nd} \alpha_{si} \beta_{si}^{T} \cdot Iv_i(n, m) \tag{4.17}$$

Where Ctr corresponds to the sum of the response output in the target area due to center processing and Sur corresponds to the sum of the response outputs in the target areas due to surround processing.

Using a $3x3$ kernel as described in Equation (4.7) we can compute Ctr and Sur to obtain the results shown in Equation (4.18) and Equation (4.19) thereby enabling the calculation of the resultant output for the target region given by J in Equation (4.15).

$$Ctr = 4\left(\beta_{c0} + 2\beta_{c1} + \beta_{c2}\right)q + 4\left(3\beta_{c2} + 2\beta_{c1}\right)p \tag{4.18}$$

$$\begin{aligned} Sur = {} & 4\beta_{s2}\left(q\right) + 4\left(\beta_{s0} + 3\beta_{s2} + 4\beta_{s1}\right)p + 8\left(\beta_{s1} + \beta_{s2}\right)q + \\ & 8(\beta_{s0} + 3\beta_{s1} + 3\beta_{s2})p + 4\left(\beta_{0s} + 2\beta_{1s} + \beta_{2s}\right)q + 4\left(3\beta_{2s} + 2\beta_{1s}\right)p \end{aligned} \tag{4.19}$$

A few more simplifications are useful for this example, we write Equation (4.18) and Equation (4.19) as shown in Equation (4.20) and Equation (4.21)

$$Ctr = Uq + Vp \tag{4.20}$$

$$Sur = Gq + Hp \tag{4.21}$$

where:

$$U = 4\left(\beta_{c0} + 2\beta_{c1} + \beta_{c2}\right)$$

$$V = 4\left(3\beta_{c2} + 2\beta_{c1}\right)$$

$$G = 4\beta_{s2} + 8\left(\beta_{s1} + \beta_{s2}\right) + 4\left(\beta_{0s} + 2\beta_{1s} + \beta_{2s}\right)$$

$$H = 4\left(\beta_{s0} + 3\beta_{s2} + 4\beta_{s1}\right) + 8(\beta_{s0} + 3\beta_{s1} + 3\beta_{s2}) + 4\left(3\beta_{2s} + 2\beta_{1s}\right)$$

We then obtain the form of the response for this example as shown in Equation (4.22)

$$J = Uq + Vp - (Gq + Hp) \tag{4.22}$$

Referring back to Figure 4.1 we would like to choose parameters to maximize the distances between the feature clusters in the response space as would be computed using Equation (4.22). Using the model shown in Figure 4.3, we look at a transition going from an area that is just in the background labeled J_B to one that includes the target area labeled J_T. We define the responses for these two areas as J_T for the target area and J_B which allows us to write the total response for the target and background area as shown in Equation (4.23)and Equation (4.24)

$$J_T = Uq_T + Vp_T - (Gq_T + Hp_T) \tag{4.23}$$

$$J_B = Uq_B + Vp_B - (Gq_B + Hp_B) \tag{4.24}$$

but $q_B = p_B = p_T$ so that we can write J_B as in Equation (4.25).

$$J_T = Up_T + Vp_T - (Gp_T + Hp_T) \tag{4.25}$$

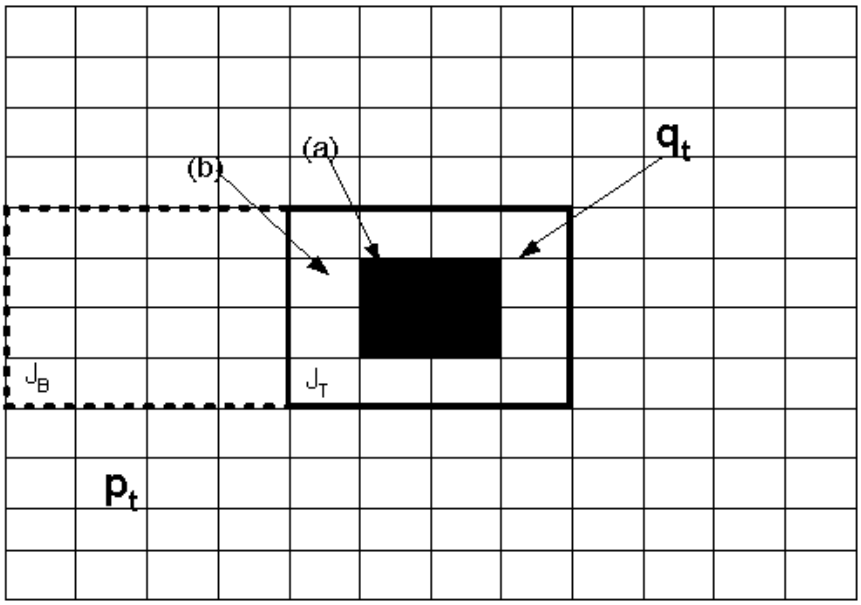

Figure 4.3: Regions used for the computation of the responses for the target and background

With these substitutions we obtain Equation (4.26) for the change in the response going from a background to a target area as shown in Figure 4.3.

$$J_T - J_B = (U - G)(q_T - p_T) \tag{4.26}$$

We would therefore like to maximize the response given in Equation (4.26) by choosing the parameters in U and G to increase the probability of detection of this target area. The parameters of interest would be σ_c and σ_s.We show the effect of the changes in these parameters on the output in Figure 4.4. It is observed that we get the maximum output as we increase σ_s and decrease σ_c subject to the constraint that $\sigma_s > \sigma_c$.

4.3.4 Response for an edge

The other problem of interest as described in Equation (4.3) was that of edge location where might also be interested in increasing our probability of detecting an edge by maximizing the response at an edge. Again, using Figure 4.3, we look a two pixels along the edge of the target. We will call the pixel in the background area (a) and the response at that pixel position R_a, similarly we will call the pixel in the target (b) and the response at that position R_b. Using a $3x3$ center surround matrix we can obtain expressions for R_a and R_b as shown in Equation (4.27) and Equation (4.28) respectively. We then look at the descriptor for the edge as the difference between R_a and R_b as shown in Equation (4.30). The response space is shown in Figure 4.5 it is also noticed that we get the maximum response when σ_s is large in comparison to σ_c subject to the same constraints on σ_s and σ_c described earlier.

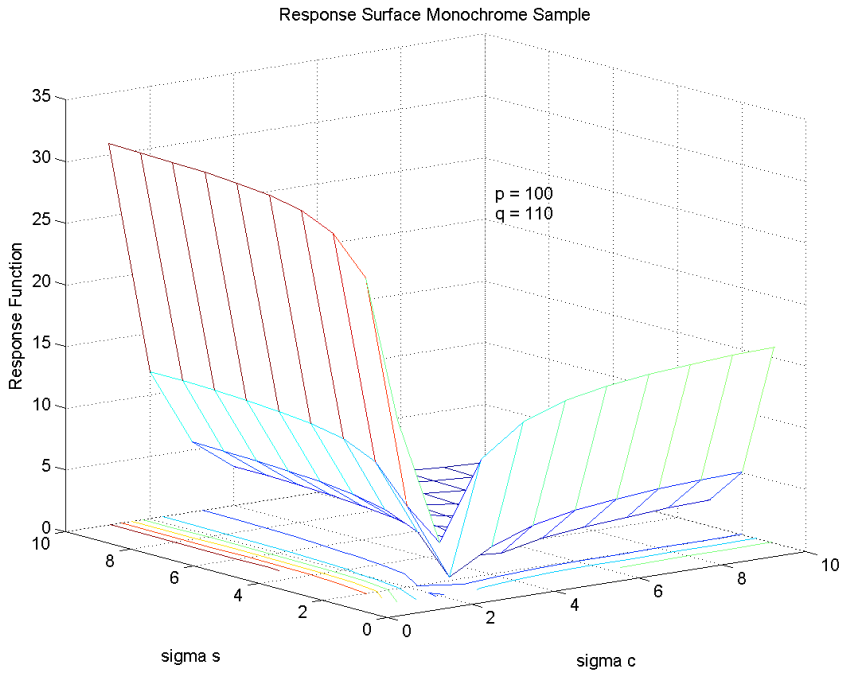

Figure 4.4: Response surface showing the effect of the sigmas

$$R_a = (\beta_{c0} + 3\beta_{c2} + 3\beta_{c1})p + (\beta_{c1} + \beta_{c2})q - (\beta_{s0} + 3\beta_{s2} + 3\beta_{s1})p - (\beta_{s1} + \beta_{s2})q \tag{4.27}$$

$$R_b = (3\beta_{c2} + 2\beta_{c1})p + (\beta_{c0} + 2\beta_{c1} + \beta_{c2})q - (3\beta_{s2} + 2\beta_{s1})p - (\beta_{s0} + 2\beta_{s1} + \beta_{s2})q \tag{4.28}$$

$$R_b - R_a = [(\beta_{c0} - \beta_{s0}) + (\beta_{c1} - \beta_{s1})]q - [(\beta_{c0} - \beta_{s0}) + (\beta_{c1} - \beta_{s1})]p \tag{4.29}$$

$$R_b - R_a = [(\beta_{c0} - \beta_{s0}) + (\beta_{c1} - \beta_{s1})](q - p) \tag{4.30}$$

We should also look at the behavior of the objective function to see if these results are reasonable

4.3.5 General Behavior of Response Functions

Given a function $z = z(x_1, x_2)$ it can be shown based on Lagrange's conditions [35] that we can infer the existence of maxima or minima of the function.

First define

$$Q_1 = \left| \frac{\partial^2 z}{\partial x_1 \partial x_2} \right| x1 = x1^* x_2 = x_2^* \tag{4.31}$$

and

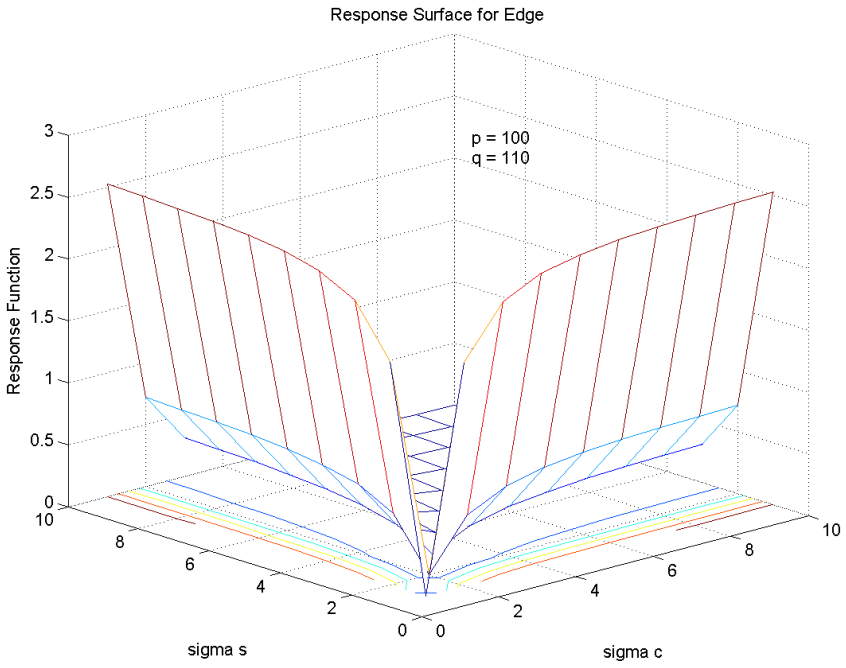

Figure 4.5: Edge descriptor response as a function of sigmas

$$Q_2 = \left| \begin{array}{cc} \frac{\partial^2 z}{\partial x_1^2} & \frac{\partial^2 z}{\partial x_1 \partial x_2} \\ \frac{\partial^2 z}{\partial x_2 \partial x_1} & \frac{\partial^2 z}{\partial x_2^2} \end{array} \right|_{x1=x1^* x_2=x_2^*} \tag{4.32}$$

where x_1^* and x_2^* are points that define maxima or minima. For a local maxima it is required that $Q_1 < 0$ *and* $Q_2 > 0$.

We can write the general response function from Equation (4.1) as in Equation (4.33).

$$R_i = \alpha_c z_c(\sigma_c) + \alpha_s z_s(\sigma_s) \tag{4.33}$$

where R_i is a linear function of α_c and α_s and thus, have no maximum for these variables as they could be chosen to produce any desired value. In most models they are assumed to be unity[13]. Computing Q_1 and Q_2 defined above we get

$$Q_1 = \frac{\partial^2 R_i}{\partial \sigma_c \partial \sigma_s} = 0 \tag{4.34}$$

$$\frac{\partial R_i}{\partial \sigma_c} = z_c'(\sigma_c) \tag{4.35}$$

$$\frac{\partial^2 R_i}{\partial \sigma_c^2} = z_c''(\sigma_c) \tag{4.36}$$

which exists since the functions z are exponentials. Similarly

$$\frac{\partial^2 R_i}{\partial \sigma_s^2} = z_s''(\sigma_s) \tag{4.37}$$

so that

$$Q_2 = z_c''(\sigma_c)\ z_s''(\sigma_s) > 0 \tag{4.38}$$

This would imply that $Q_1 = 0$ and $Q_2 > 0$ which implies that for these functions maximas would tend to occur on saddle points or on boundaries.

The general behavior of this $Class\ I$ response and the effect of its parameters for the monochromatic image model are shown in Figure.4.6 through Figure 4.8. Figure 4.6 shows the general character of the response as we vary σ_c while Figure 4.7 and Figure 4.8 shows the effect of σ_s and q. As expected, the maximas occur on the boundaries or on saddle points. The relative image values affect the general shape of the decision surface.

At this point we will now examine the behavior for the response functions for the Class I and Class II responses for a simple color image as presented in Figure 4.9 where we have a background region with pixel values $p_r,\ p_g$, and p_b while the target area has values $q_r,\ q_g$, and q_b. Using Equation (4.1) and Table 4.1 we define the general $Class\ I$ responses in Equation (4.39) through Equation (4.41).

$$R_1(m,n) = \alpha_c\beta_{c1}(m,n) * IR(m,n) - \alpha_s\beta_{s1}(m,n) * IR(m,n) \tag{4.39}$$

$$R_2(m,n) = \alpha_c\beta_{c2}(m,n) * IG(m,n) - \alpha_s\beta 2(m,n) * IG(m,n) \tag{4.40}$$

$$R_3(m,n) = \alpha_c\beta_{c3}(m,n) * IB(m,n) - \alpha_s\beta_{s3}(m,n) * IB(m,n) \tag{4.41}$$

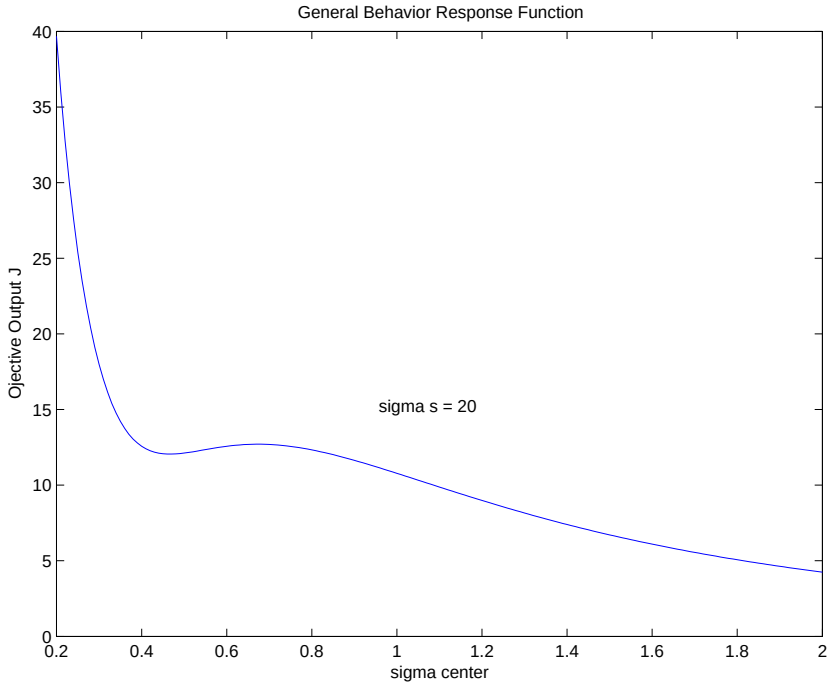

Figure 4.6: General Class I response as it varies with σ_c

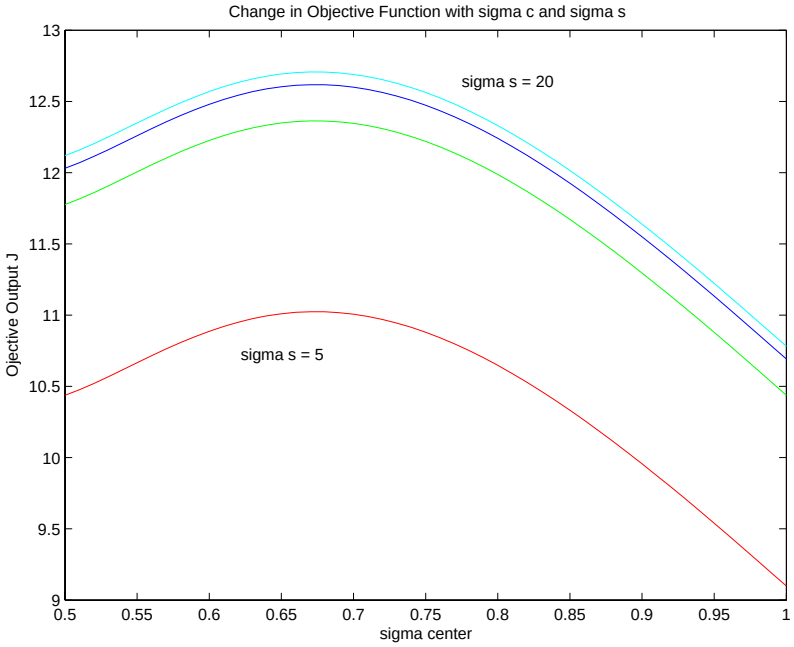

Figure 4.7: Effects of sigma c and sigma s on response

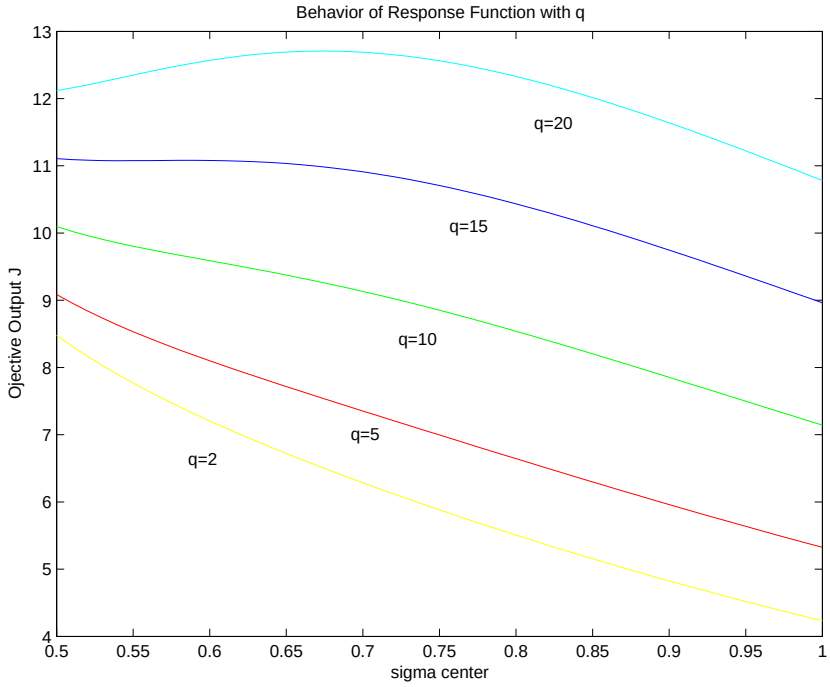

Figure 4.8: Class I Response for relative values of q with p=10 from the model image

The general $Class\ II$ responses are given in Equations (4.42) through Equations (4.44).

$$R_4(m,n) = \alpha_c\beta_{c1}(m,n) * IR(m,n) - \alpha_s\beta_{s1}(m,n) * (IR(m,n) - IG(m,n)) \quad (4.42)$$

$$R_5(m,n) = \alpha_c\beta_{c1}(m,n) * IG(m,n) - \alpha_s\beta_{s1}(m,n) * (IR(m,n) - IG(m,n)) \quad (4.43)$$

$$R_6(m,n) = \alpha_c\beta_{c1}(m,n) * IB(m,n) - \alpha_s\beta_{s1}(m,n) * (IR(m,n) + IG(m,n) - IB(m,n)) \quad (4.44)$$

IR, IG, and IB are the red, green and blue image planes respectively. We then define the response functions for the responses to the simple color image as J_1 through J_6 corresponding to the general responses R_1 through R_6. The details of the computations are presented in Appendix A.

The response surfaces for responses J_1 through J_6 with varying σ with a change from background to target such that $p_r = 140,\ q_r = 90\ p_g = 80,\ q_g = 59,\ p_b = 60,\ q_b = 51$ are shown in Figure 4.10 through Figure 4.15. We observe that we maximize the responses when we have σ_s large and σ_c small. We also observe that the difference as represented by the magnitude of the responses are greater for the $Class\ II$ responses than for the $Class\ I$ responses indicating that they are more sensitive to changes in color.

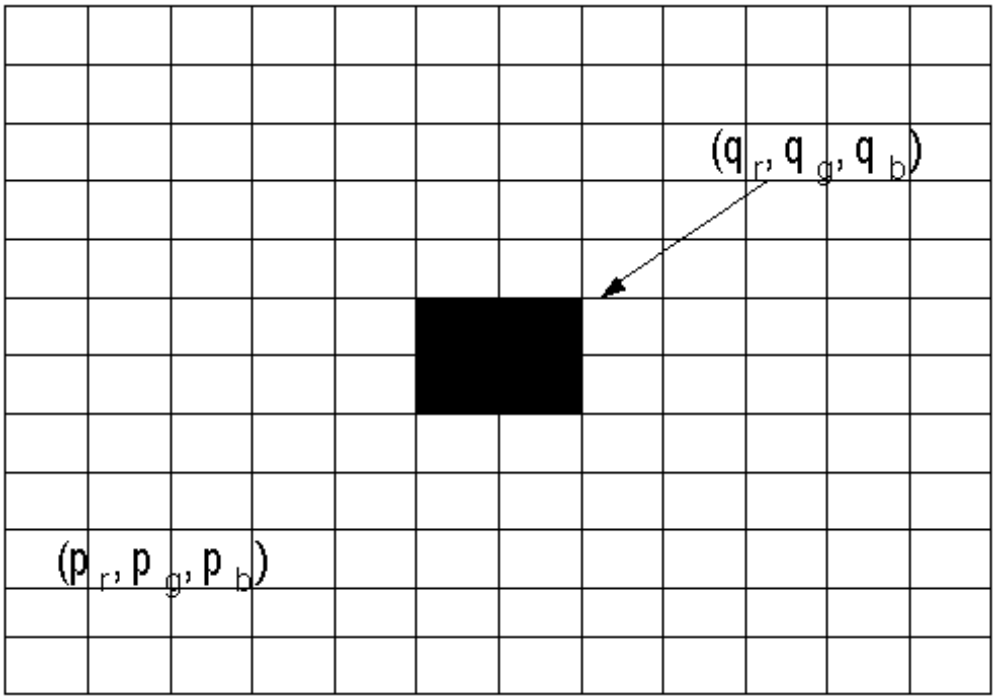

Figure 4.9: Color sample output picture pr, pg, pb are background values, qr, qg and qb are target values

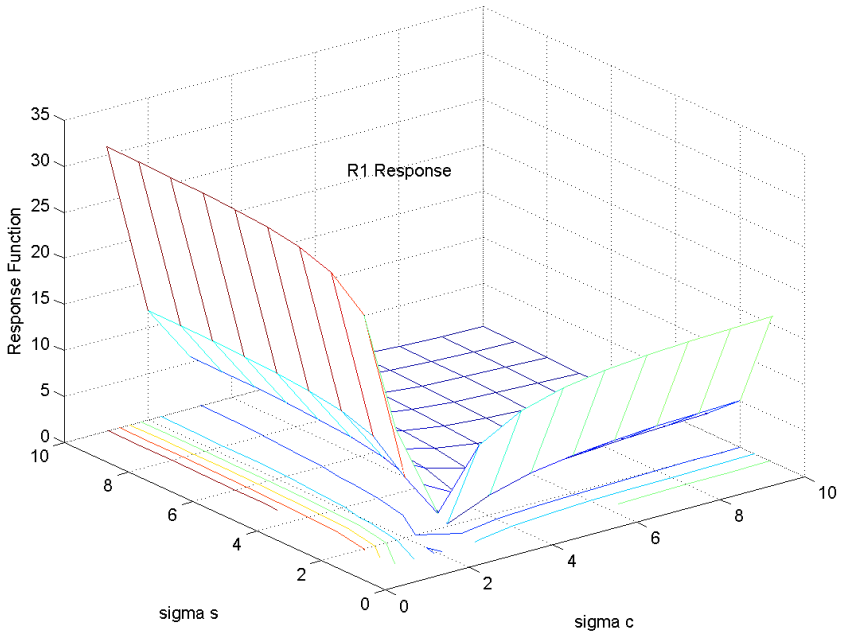

Figure 4.10: Response for output J1 as a function of σ

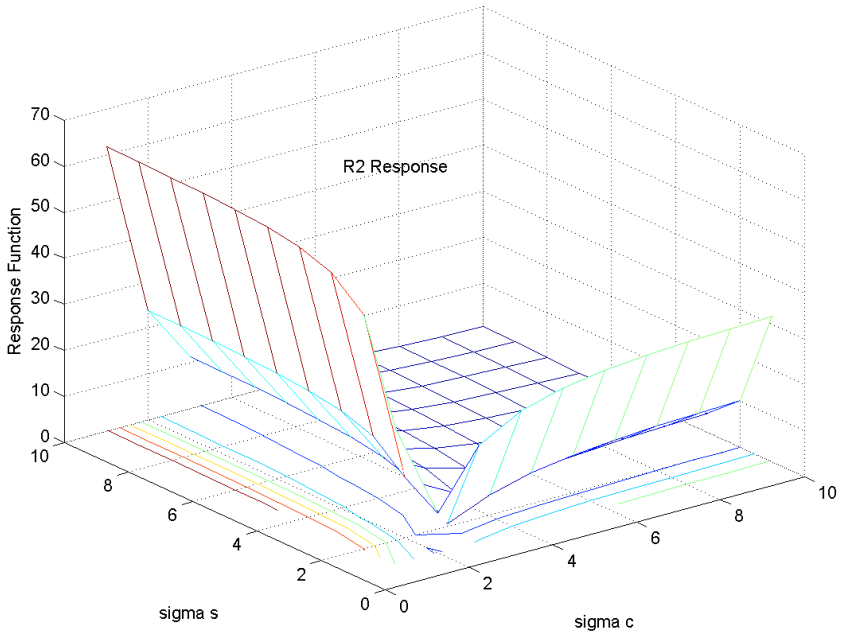

Figure 4.11: Response of function J2 as a function of σ

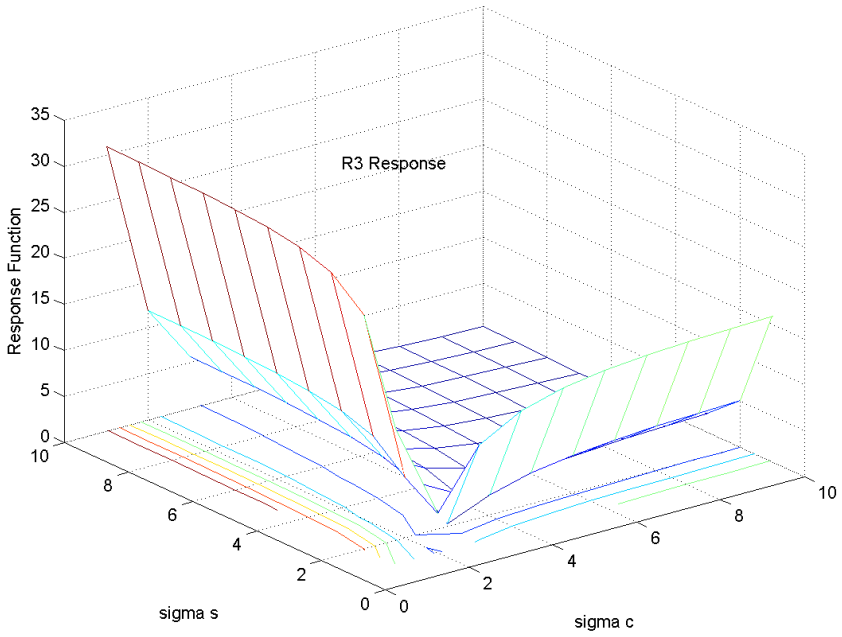

Figure 4.12: Response of function J3 as function of σ

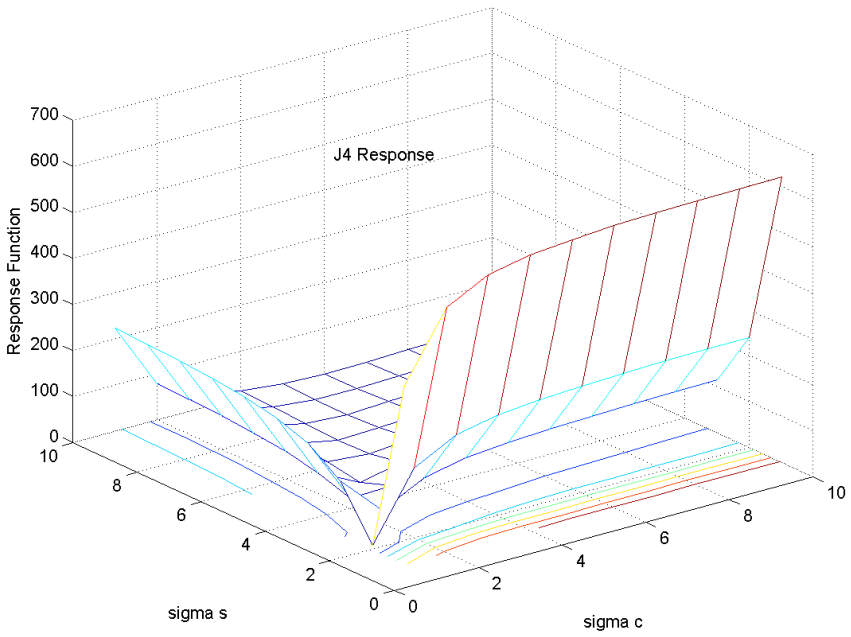

Figure 4.13: Response of function J4 as a function of σ

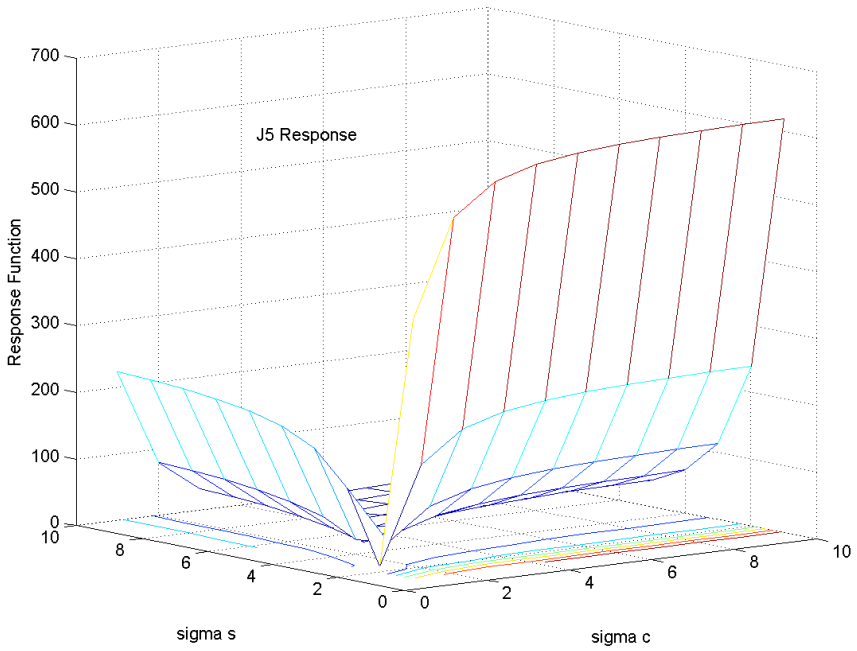

Figure 4.14: Response of function J5 as a function of σ

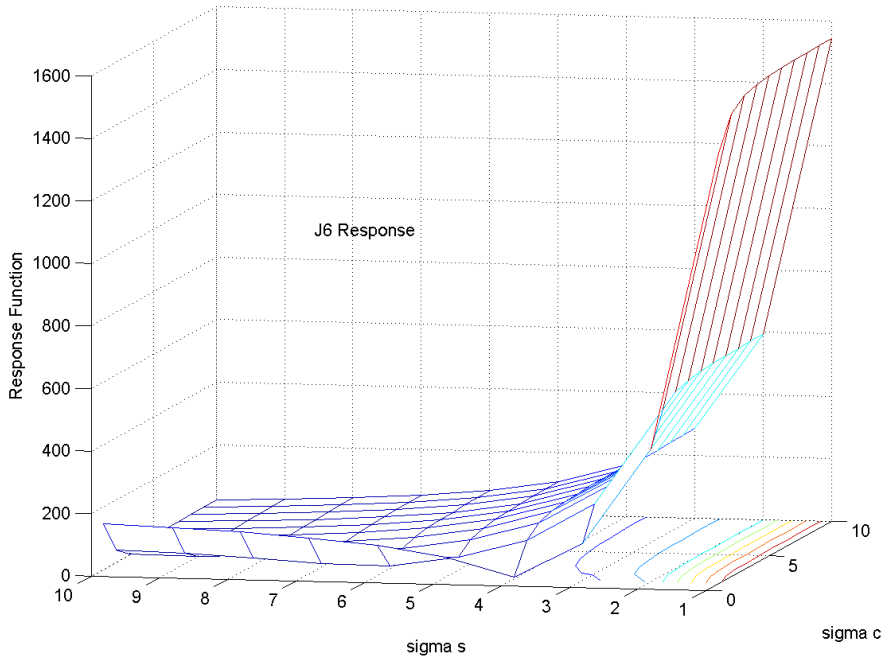

Figure 4.15: Response of function J6 as a function of σ

Table 4.2: Class I output response with parameters as shown

Parameters	p	J
$r = 0.1$	10	7.46
$\sigma_c = 1$	20	14.93
$\sigma_s = 10$	30	22.38
	40	29.85

4.4 Comparison with Biological Experiments

A relevant question at this point is whether or not these responses are representative of the human visual response? Since we are extracting the approaches described from models of the HVS we now compare the results obtained from experiments conducted on the primate visual system. First, we show in Figure 4.16, experimental data for the threshold intensity at increasing ambient light levels for an image as in Figure 4.2 from Wandell [26]. We then compute the response using Equation (4.22) for our model where $q = (1 + r)p$. The results are and shown in Table 4.2 and Figure 4.17 respectively. These plots both show a linear response as the background intensity is changed. This implies a similar behavior to the biological system.

Additionally we could look at the response across an edge such as that shown in Figure 4.18. An image representation of the summed output responses (R_1 through R_6) computed using Equation (4.39) to Equation (4.44) is shown in Figure 4.19 as a pseudo colored image. A profile of the response along the line shown in Figure 4.19 for the colored edge in Figure 4.18 is shown in Figure 4.21. The response across a similar edge is shown in Figure 4.20 from work done by Enroth-Cugell et. al. [36]. Here, they looked at measurements of the firing rate of a ganglion cell as an edge is passed through its field of view, they are also seen to be similar in form to that illustrated in Figure 4.21. These two examples illustrate that the receptive field formulation

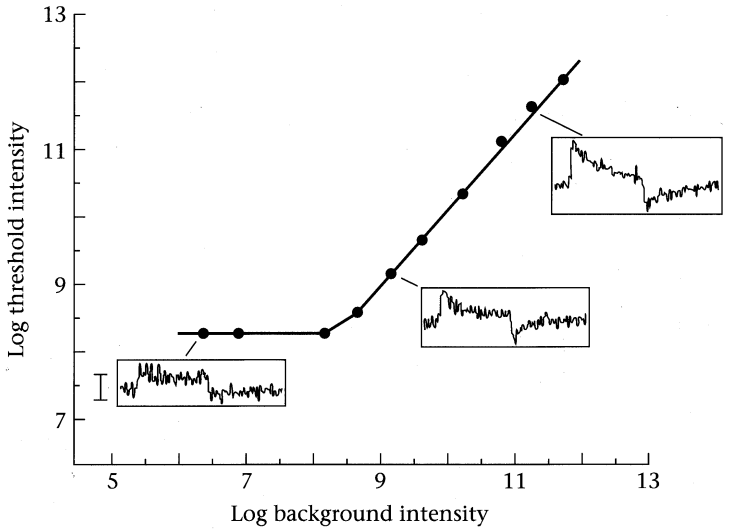

Figure 4.16: Experimental data showing threshold intensity as a function of background intensity [26]

presented here demonstrates some of the of the performance of the primate visual system at least at the level of the response of the ganglion cells. We also show in Figure 4.22 the relative characteristics of the Class I and Class II responses. It is observed that the *Class I* output is driven mostly by the edge while the *Class II* responds more to the change in color (Class I magnitude offset so that the plots have the same relative magnitude).

4.5 Summary

We have attempted in this chapter to examine the behaviors of the computational structures proposed for use to extract features that will be useful for classification in machine vision applications. We have shown that we are able to select σ_c and σ_s that

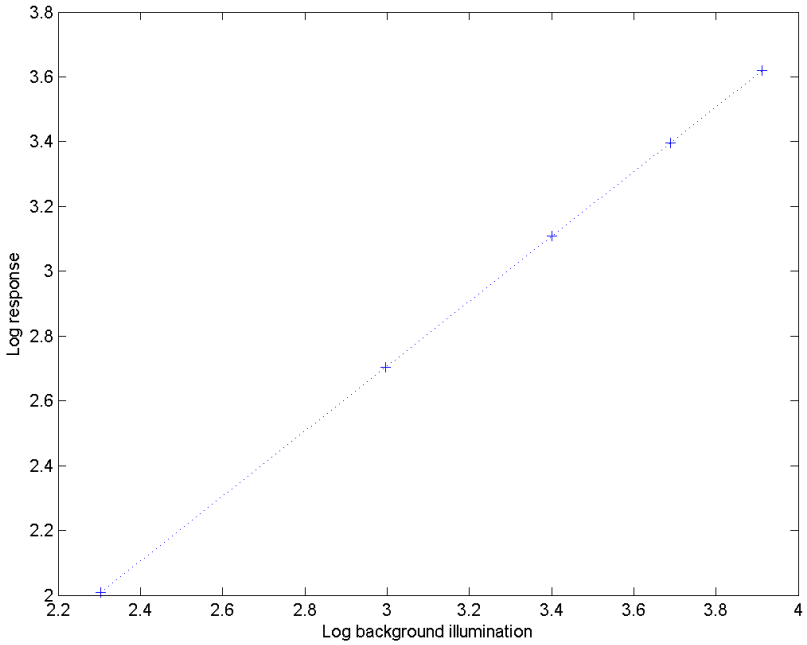

Figure 4.17: Responses for a constant ratio of target intensity to the background

Figure 4.18: A sample edge image using a blood background

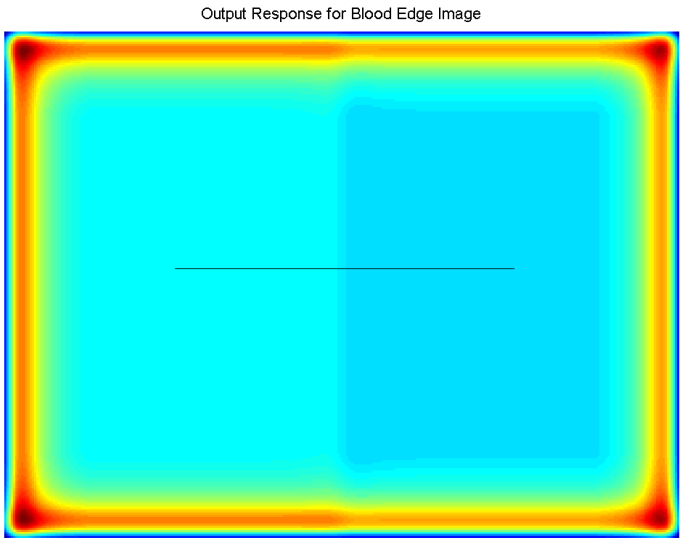

Figure 4.19: The output response for the blood edge image

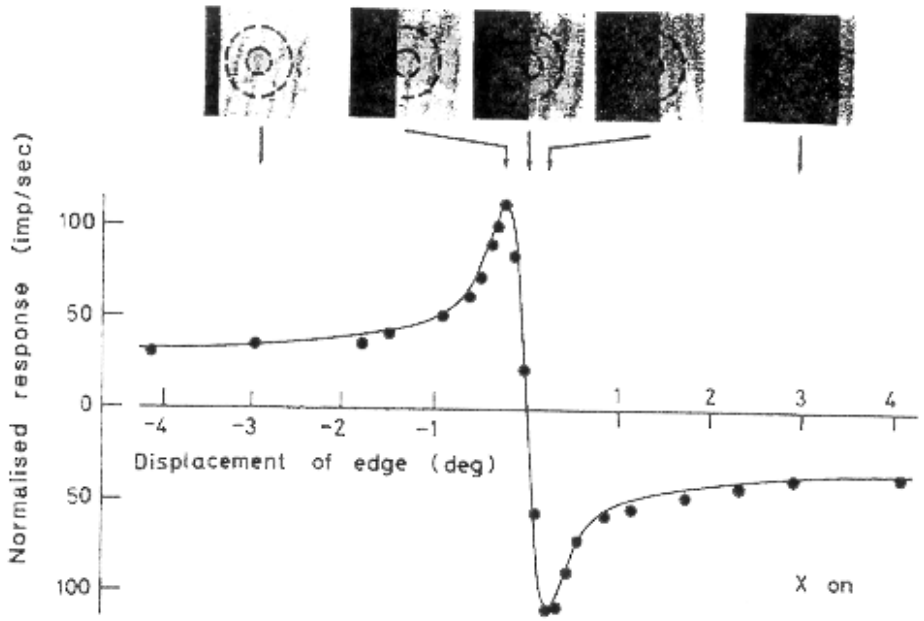

Figure 4.20: Response to a ganlion X cell response to an edge

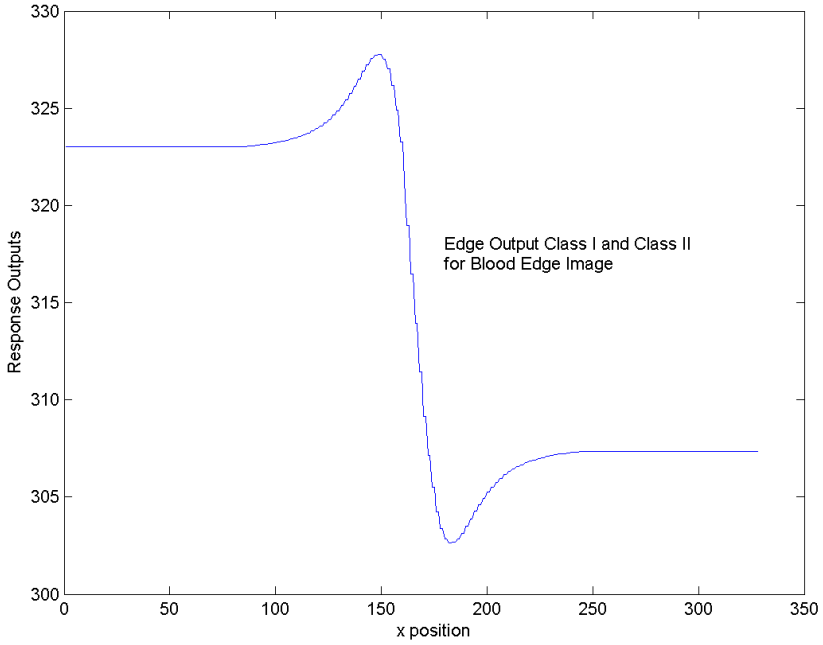

Figure 4.21: Sample edge output computed for blood edge image

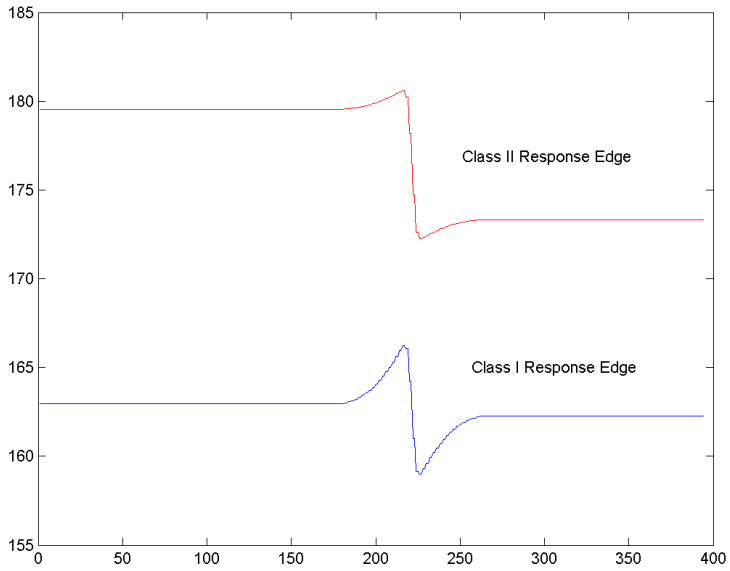

Figure 4.22: Individual Class I and Class II responses

would allow us to maximize the difference between features representing objects of interest from the background. It is known that the human eye adjusts the relative sizes of the receptive fields based on the illumination conditions in the image with typical ratios being 1 to 10. This analysis describes a region within which those adjustments would take place. Additionally it is observed that the $Class\ I$ structures do not respond as strongly to color changes as the $Class\ II$ structures while the $Class\ I$ structures respond more strongly to edges. This implies that for applications requiring fine color discrimination the $Class\ II$ responses would provide more utility. We also showed that the representations described demonstrate some of the characteristics of the human visual system at the level of the ganglion cell responses. In Chapter 5 we will look at the development of some practical applications using the techniques. The approach described provides a very straight forward implementation as will be demonstrated with the example applications.

Chapter 5

Testing Contrast Feature Approaches

The general approach as described in the previous chapters was to extract and exploit image features derived from the current state of knowledge about the human visual system. Specific features were identified in terms of the responses of different mechanisms in the eye brain system and approaches towards their calculation and utilization explored. This chapter presents specific applications of these approaches and their results. We first look at data using simulated images and then using data from real problems.

5.1 Testing with Simulated and Real Data

5.1.1 Simulated Data

We have developed techniques that can be used for identifying regions of interest in an image based on proposed biological models. We use these models in simulation to understand the expected behavior for natural images. The images are shown in Figure 5.1(a), (b) and (c) and represent simulated breast 'butterflies'(a particular cut of meat) with blood, fan bone, and cartilage respectively, which are all features of interest in this problem area. These images were generated using the techniques

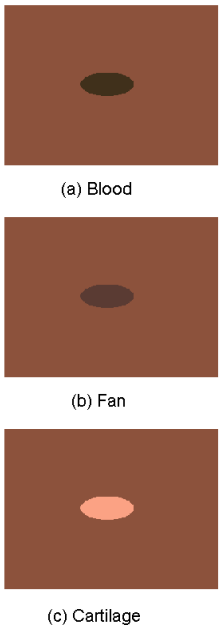

Figure 5.1: Sample prototype images on a meat background (a) blood, (b) fan bone, (c) cartilage

and tools presented in Appendix B and Appendix C using irradiance data for the actual items of interest. It can be observed that the blood and fan bone are similar in appearance and could be difficult to distinguish if viewed at fast rates.

The results from the analysis in the previous chapter for the *Class I* responses indicate that we obtained the maximum separation in the response space by choosing the minimum σ_c and the maximum σ_s. The sample images were processed using Equation (4.1) with $\sigma_c = 1$ and $\sigma_s = 10$. The intermediate images generated in the processing steps are shown in Figure 5.2 for the *Class I* responses for the prototype blood image where (a) is the original image, (b) the center output, (c) the surround

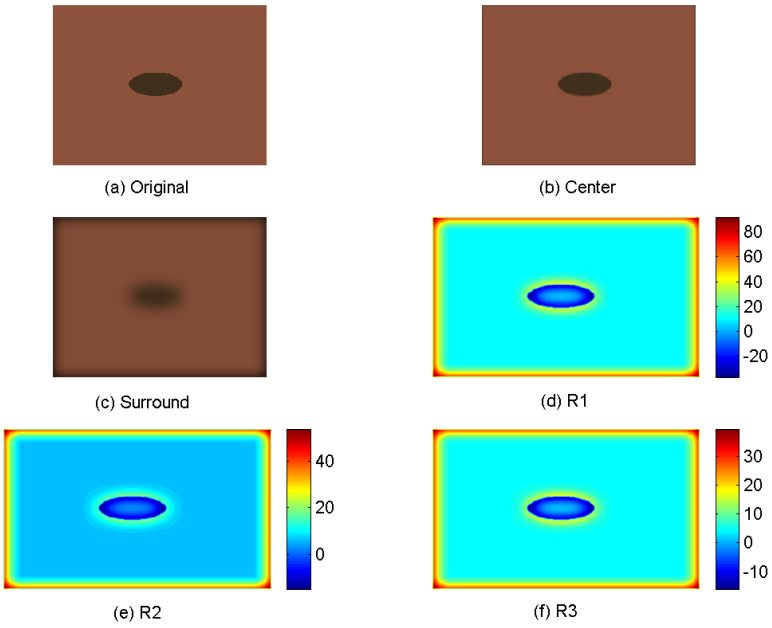

Figure 5.2: Sequence of images used in processing (a) Original, (b) Center, (c) Surround, (d) R1, (e) R2, (f) R3

output, (d) response 1 (R1), (e) response 2 (R2), and, (f) response 3 (R3). In Figure 5.3 we show the image representation of the *Class II* output responses for the same image.

Looking at the output images presented in Figure 5.2 and Figure 5.3 we are able to see the larger magnitudes of the color transitions between the *Class I* and the *Class II* responses.

The resulting data for these three cases under *Class I* and *Class II* processing are presented in Table 5.1 and Figure 5.4 and Figure 5.5 respectively. It should be noticed that even though the background is the same in all three images the responses for

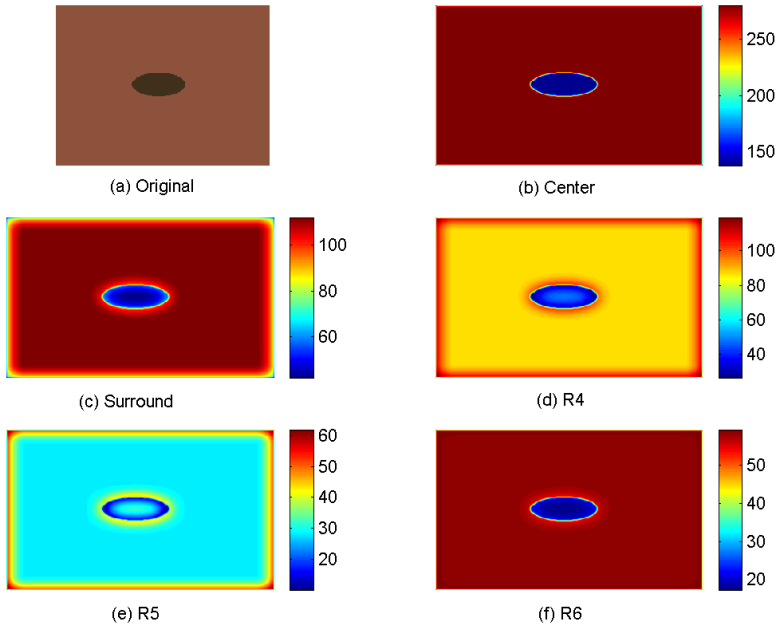

Figure 5.3: Intermediate outputs for the Class II responses (a) Original, (b) Center, (c) Surround, (d) R4, (e) R5, (f) R6

Table 5.1: Class I and Class II prototype responses

Class I			**Class II**		
	Feature	**Background**		**Feature**	**Background**
Blood					
R1	-13.64	12.21	R4	40.04	87.15
R2	-4.57	6.99	R5	22.67	29.24
R3	-5.99	5.24	R6	35.70	61.93
Fan					
R1	-5.19	11.82	R4	55.56	86.94
R2	-1.00	6.82	R5	23.68	29.00
R3	1.812	4.87	R6	64.70	62.14
Cartilage					
R1	47.12	9.35	R4	174.04	86.27
R2	32.47	5.25	R5	86.05	28.20
R3	28.13	3.64	R6	157.84	62.70

the background regions are different as the response is also dependent on the feature of interest (contrast between them). Looking at the outputs in the response spaces shown in Figure 5.4 and Figure 5.5, it is observed that it would be a fairly simple affair to develop a classifier for these different kinds of defects.

Again it can be observed that the *Class II* responses are in general stronger than those for *Class I* possibly signifying the ability to more accurately discern finer color differences than the Class I outputs. We infer this from the fact that the Class II response difference in going from the background to the feature of interest is almost double that for the Class I. Using Table 5.1 we find, for blood it is 54.3 versus 30.46, and for fan bone it is 31.93 versus 18.97. In both cases presented here however, it would be possible to differentiate the blood from the fan bone even though from the images they appear close in color.

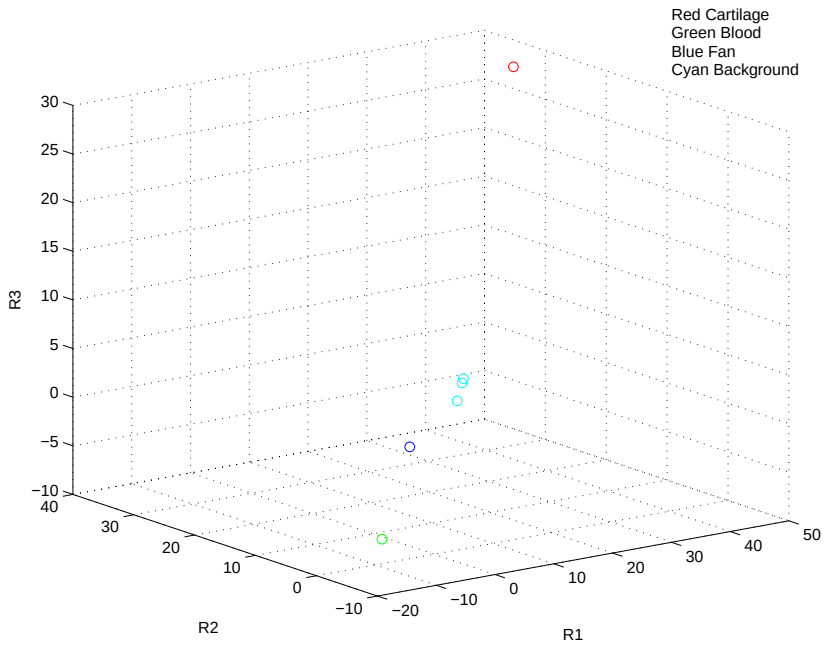

Figure 5.4: Distribution of response outputs for blood, fan and cartilage

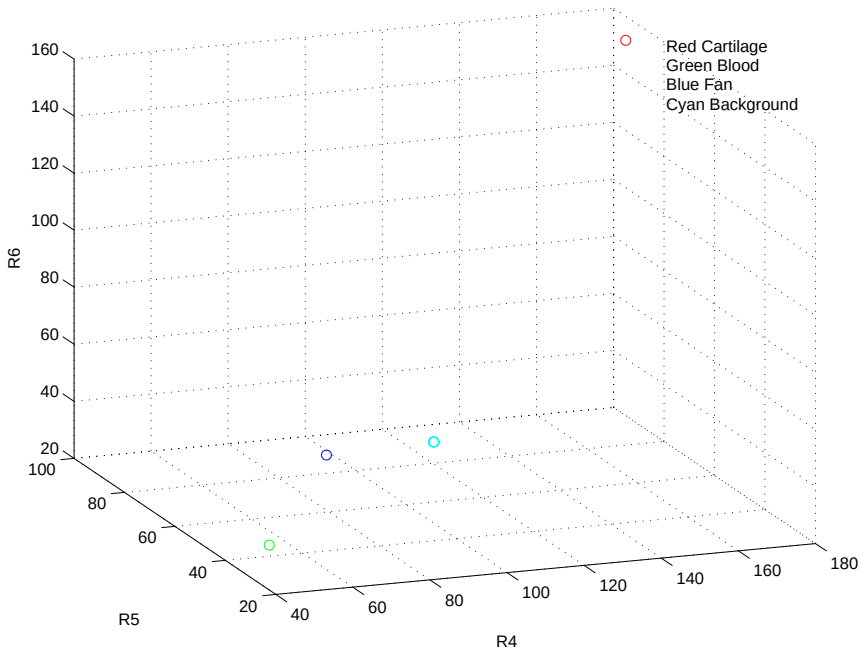

Figure 5.5: Responses for Class II type Outputs

5.1.2 Responses for Real Images

Next we will look at the responses that would be obtained using real artifacts. The first to be examined will be chicken 'butterflies' with fan bone defects. A sample fan bone is recognized as the fan shaped area in the lower right corner of Figure 5.6. More detail on the application will be given later in Section 5.2.3 Example Applications. The goal is to develop a representation that allows us to recognize fan bones but would still be robust in light of the expected natural variability along with the expected deviations in camera and imaging parameters etc. We will first look at sample data that illustrates the effects due to changes in imaging parameters. These changes could simulate the differences that would be seen by different human observers conducting the same task. One approach to address this would be to try and adjust the lighting and imaging parameters to obtain the same appearance. This is a difficult proposition in most practical applications because of the naturally occurring changes in the environment; it would also be difficult to accommodate these changes in real time.

A sample fan bone image taken with a 3-CCD camera at f-stop 4.0 is presented in Figure 5.6. Figure 5.7 shows the same part imaged at f-stop 5.6. We then look at the 'raw'(original RGB) and 'processed' (after filtering with Class II and Class I operators) output region values for the feature of interest (fan bone) to determine which representation would be more beneficial for processing and analysis.

Sample scatter data for these regions are presented in Figure 5.8. We will next look at measures for comparing the representations using the scatter data. This data is presented in Table 5.2 and Table 5.3 and the general idea is that smaller variances and

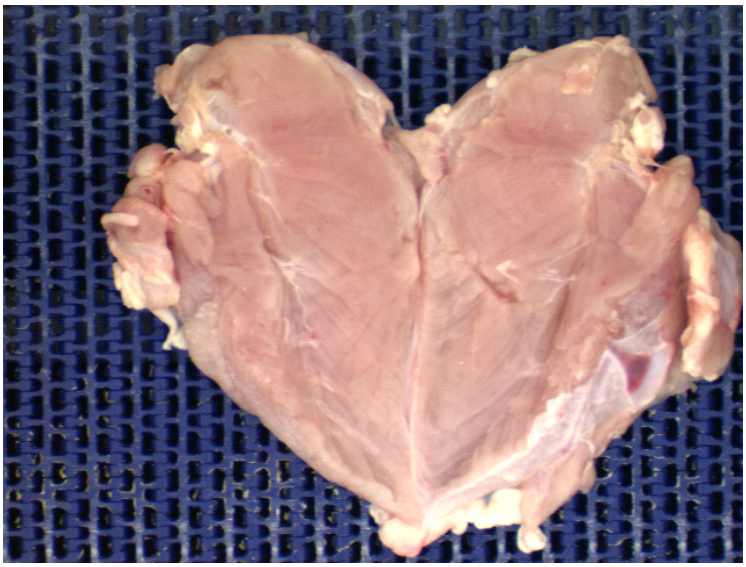

Figure 5.6: Fanbone image taken with 3-CCD camera at f-stop 4.

consistent means will result in a more stable representation for algorithms that would be used to classify these entities. Several things can be observed; first, the variances and the eigenvalues of the processed data are, in general, smaller indicating a more compact representation. In addition there also appears to be smaller changes in the means of the feature values indicating a more stable representation for algorithms to do the classification.

We will now look at what happens when we change cameras. The image shown in Figure 5.9 is of the same part but taken with a single CCD digital camera. Because of the Bayer filter on the image sensor we would expect this image in general to be more noisy in nature. The covariance matrix and mean values for fan bone and background regions in this image are also presented in Table 5.4. Again we observe that even though there are significant changes in the raw image representations the

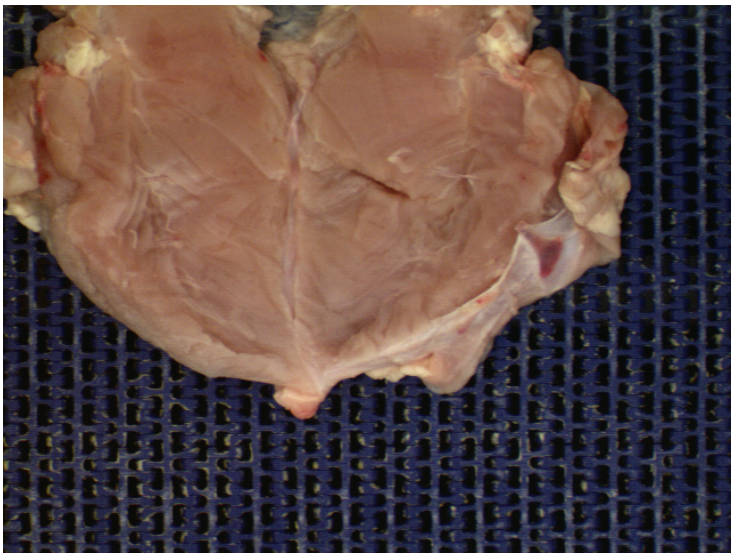

Figure 5.7: Same part as in the previous image taken at f-stop 5.6

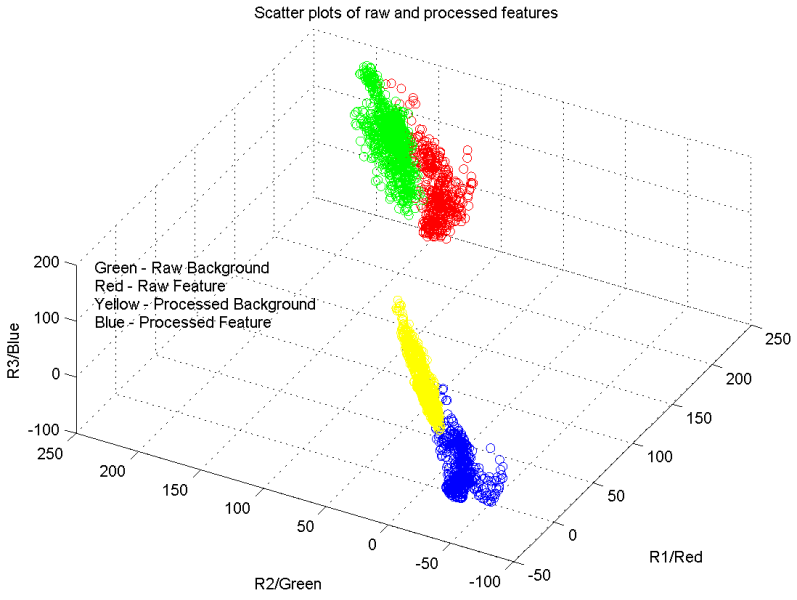

Figure 5.8: Scatter plots for feature and background in Figure 5.6

Table 5.2: Processed and raw data for image at f4.0

Processed					
Covariance			Eigenvectors		
254.53	178.02	139.79	-0.1040	0.7899	0.6043
178.02	229.04	177.73	0.6952	-0.3768	0.6122
139.79	177.73	173.85	-0.7113	-0.7113	0.5100
Mean			Eigenvalues		
-11.92	-33.53	-25.60	20.57	84.01	552.84
Raw					
Covariance			Eigenvectors		
442.64	341.40	270.61	-0.1417	0.7686	0.6239
341.40	391.92	295.54	0.7124	-0.3584	0.6033
270.61	295.54	289.40	-0.6873	-0.5299	0.4967
Mean			Eigenvalues		
171.74	103.28	117.62	38.87	96.83	988.26

Table 5.3: Processed and raw data for image at f5.6

Processed					
Covariance			Eigenvectors		
116.88	85.47	67.20	-0.0944	0.8506	0.5172
85.47	134.86	110.52	0.6873	-0.3202	0.6520
67.20	110.52	103.97	-0.7202	-0.4171	0.5544
Mean			Eigenvalues		
-9.28	-27.55	-22.63	7.3083	51.76	296.64
Raw					
Covariance			Eigenvectors		
211.83	174.79	131.42	-0.2643	0.7773	0.5709
174.79	238.47	164.79	0.7353	-0.2206	0.6408
131.42	164.79	174.33	-0.6241	-0.5891	0.5133
Mean			Eigenvalues		
126.27	70.55	69.27	35.80	62.63	526.20

processed image parameters remain relatively close to those computed before implying that classifications based on these features would be more likely to be robust. This conclusion is also supported by a statistical analysis of the changes in the means. This was done by testing the hypothesis that there was no change in the mean values of the feature using a MANOVA (Multivariate Analysis of Variance) analysis. The results are presented in Table 5.5 and indicate that even though there are changes in the means under the different conditions, the changes of the mean feature values are much less for the processed than the unprocessed image data. For the means to be the same the F value computed would have to be less than the test statistic, which it is not for either case. It is observed that it is significantly higher for the raw case however implying that the changes in the means are larger here than under Class I processing. This conclusion is also supported if we look at the Mahalanobis distances [37] (this takes into account the variance of the data) between the clusters for the raw and processed data as presented in Table 5.6 where we see that the distances between the clusters are significantly less for the processed than the raw data.

5.1.3 Real Data Class II Responses

The other outputs that are of interest to us are the Class II outputs. Scatter plots for the same data presented earlier for the Class I outputs are now presented for the Class II outputs in Figure 5.10. The corresponding statistics are presented in Table 5.7. It is observed that the variances here are comparable to that for the raw color data and that the means have about the same variability as shown by the MANOVA test result in Table 5.8. The magnitude of the responses are much higher than for the Class I responses however matching the results of the simulated images presented earlier. In

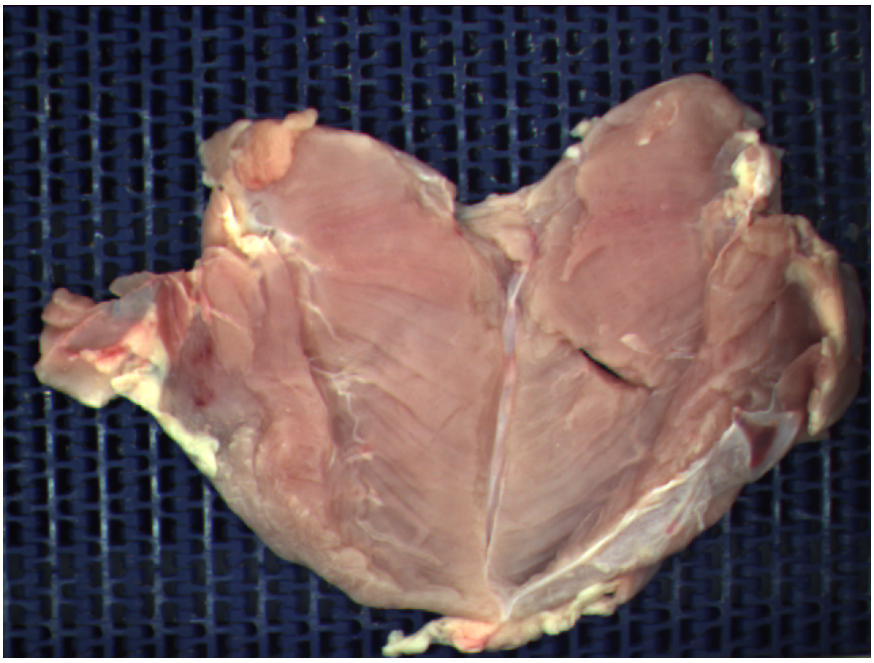

Figure 5.9: Same part taken with a single chip digital camera

Table 5.4: Processed and raw data for single chip camera

Processed					
Covariance			Eigenvectors		
213.26	113.20	117.80	0.0440	0.7087	0.7042
113.20	111.68	104.59	0.6807	-0.5372	0.4981
117.80	104.59	111.08	-0.7312	-0.4574	0.5060
Mean			Eigenvalues		
-7.59	-21.97	-17.30	6.64	51.41	377.97
Raw					
Covariance			Eigenvectors		
477.35	331.71	288.06	-0.1251	0.7202	0.6824
331.71	302.83	257.57	0.7434	-0.3875	0.5452
288.06	257.57	255.87	-0.6571	-0.5754	0.4869
Mean			Eigenvalues		
89.33	52.55	55.68	19.33	68.76	947.97

Table 5.5: MANOVA results for raw and processed data

MANOVA Test	Test Statistic	F
Processed	0.8314	45.42
Raw	0.1276	845.18

Table 5.6: Mahalanobis distances between clusters for f4.0, f5.6 and single chip

Mahalanobis	D1	D2	D3
Processed	6.96	8.85	3.84
Raw	23.54	27.72	7.4

order to examine this effect we used the image shown in Figure.5.11 This image is a possible problem as it has a bruise on the lower left of the butterfly in approximately the same position as the fan bone on the right side. The visual properties of the fan bone and bruise are also similar. The algorithms used, utilize the position of the potential fan bone regions as part of 'noise filtering' so that this area could be falsely classified as a fan bone. It would be desirable for our preprocessing steps to not present this area as a candidate fan bone region.

The scatter plots of the data for the bruise and fan bone regions in terms of Class I and Class II responses are shown in Figure 5.12. It is observed that there is much cleaner separation for the Class II output case compared to the Class I. The MANOVA results also support this conclusion as shown in Table 5.9 and Table 5.10.

5.2 Implementation of the Technique

5.2.1 Summary of the Process and Implementation

To this point we have described an approach towards the development of machine vision algorithms utilizing models based on the human visual system (HVS). We

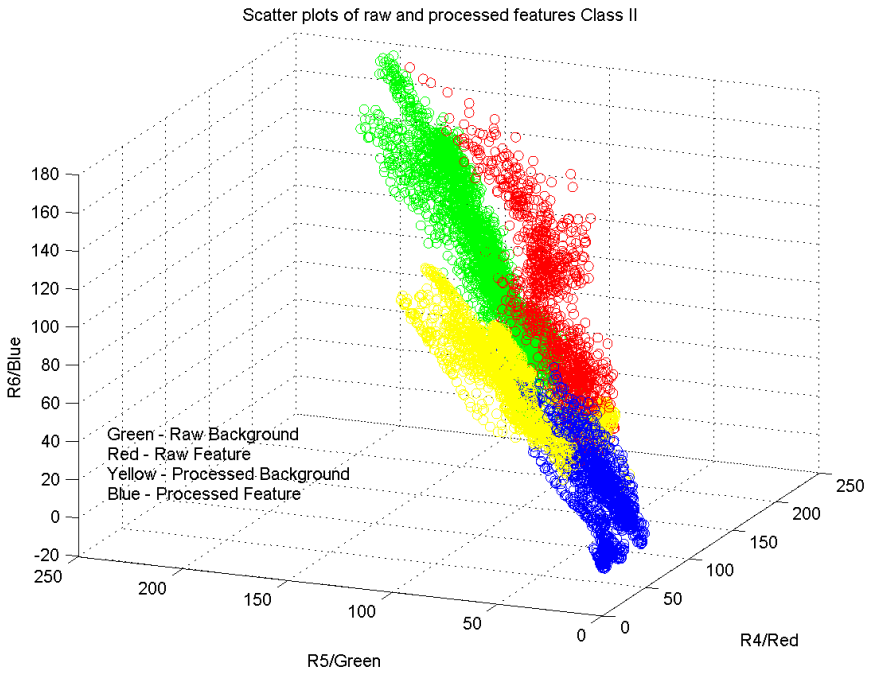

Figure 5.10: Scatter plots for Class II outputs

Table 5.7: Processed data for Class II (f4.0, f5.6, SingleChip)

Class II Processed f4.0					
Covariance			Eigenvectors		
484.11	409.53	144.86	-0.437	-0.588	0.681
409.53	448.22	199.23	0.693	0.263	0.672
144.86	199.23	155.26	-0.574	0.765	0.292
Mean			Eigenvalues		
126.28	58.29	20.77	24.97	112.41	950.21
Class II Processed f5.6					
Covariance			Eigenvectors		
255.16	231.2	66.50	-0.487	-0.584	0.649
231.2	271.67	120.91	0.679	0.213	0.702
66.50	120.81	98.36	-0.549	0.783	0.293
Mean			Eigenvalues		
90.20	35.00	14.31	7.993	81.653	535.53
Class II Processed Single Chip					
Covariance			Eigenvectors		
397.54	288.44	193.15	-0.180	0.677	0.714
288.44	255.12	177.43	0.703	-0.419	0.574
193.15	177.43	138.97	-0.688	-0.605	0.400
Mean			Eigenvalues		
68.343	32.185	4.485	7.805	45.92	737.71

Table 5.8: MANOVA Test for Class II

MANOVA Test	Test Statistic	F
Processed Class II	0.1793	639.373

Table 5.9: Class I test results for bruise and bone

MANOVA Test	Test Statistic	F
Processed Class I	0.5221	204.4

Table 5.10: Class II test results for bruise and bone

MANOVA Test	Test Statistic	F
Processed Class II	0.03884	5527.1

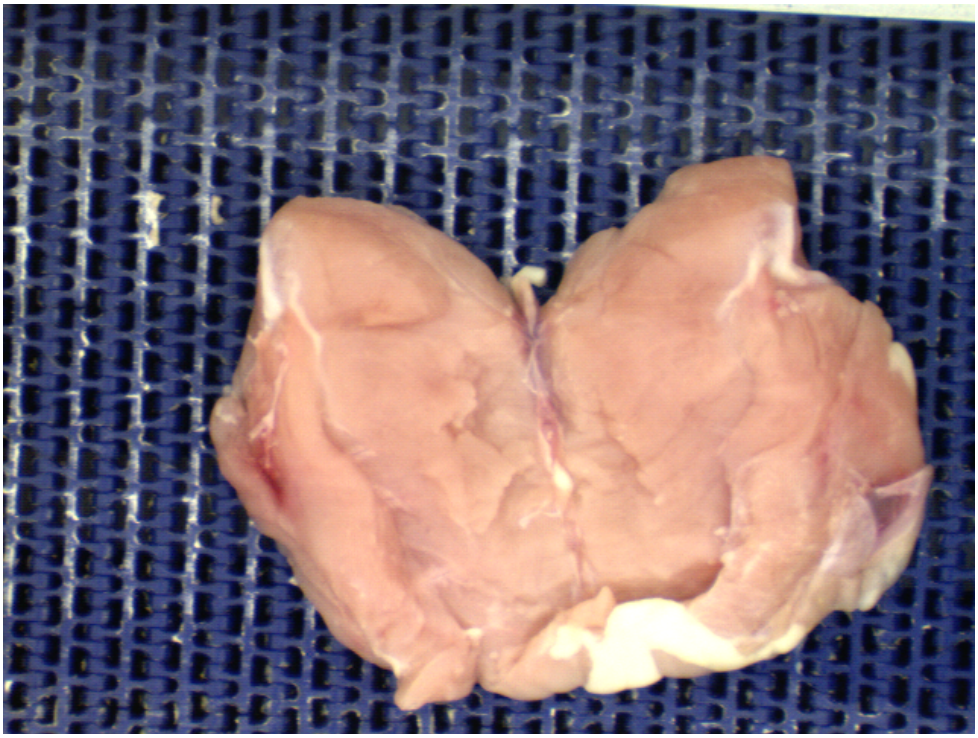

Figure 5.11: Part with both bruise and fanbone

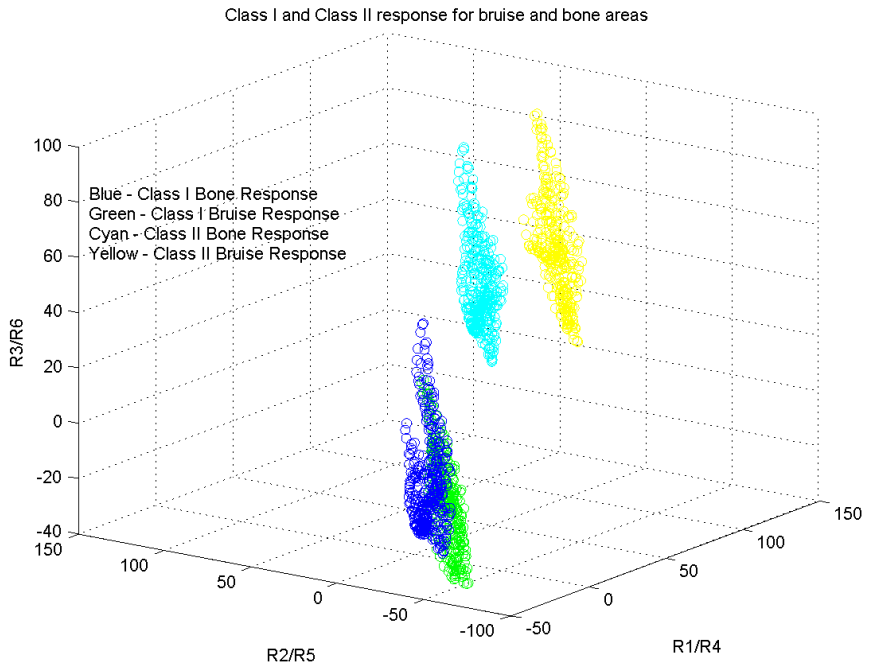

Figure 5.12: Scatter plots for Class I and Class II bone and bruise

will now summarize the approach and its implementation as it would be applied to problems of interest and the steps necessary to implement solutions.

The overall procedure is shown in the flowchart presented in Figure 5.13. The process assumes that images acquired could adequately depict the scenes of interest. We assume that the elements of interest in the scene can be identified by people that normally conduct these tasks. The first step then is for the expert to identify defect areas and background areas respectively for several sample images. Next, we a Gaussian filter that minimizes σ_c and maximizes σ_s. The next step computes $Class\ I$ and $Class\ II$ outputs for the regions identified (target and background) and examines their clustering in the $Class\ I$ and $Class\ II$ response spaces. We then apply the spherical conversion operation described in Section 5.1 and determine the cluster boundaries that could be used for classification. The representation that provides the desired results are then chosen for implementation and testing.

The processed output space is a convolution with a filter that is a linear combination of Gaussians; but the real effect is the difference between a smoothed (or averaged version of the image) and the original (or non-filtered image in the limit) images. Because of processing speed requirements, we implemented this by generating the output image as the difference between the original and an average value for the whole image. This average image was generated by computing the mode of the pixel values in the image for each image plane. This was the approach taken in the example applications to be described in the upcoming sections.

In the applications that typically revolve around machine vision, speed, (time to complete the analysis of an image) is usually a significant consideration. The allowable time for most applications are usually on the order of seconds or fractions of a second.

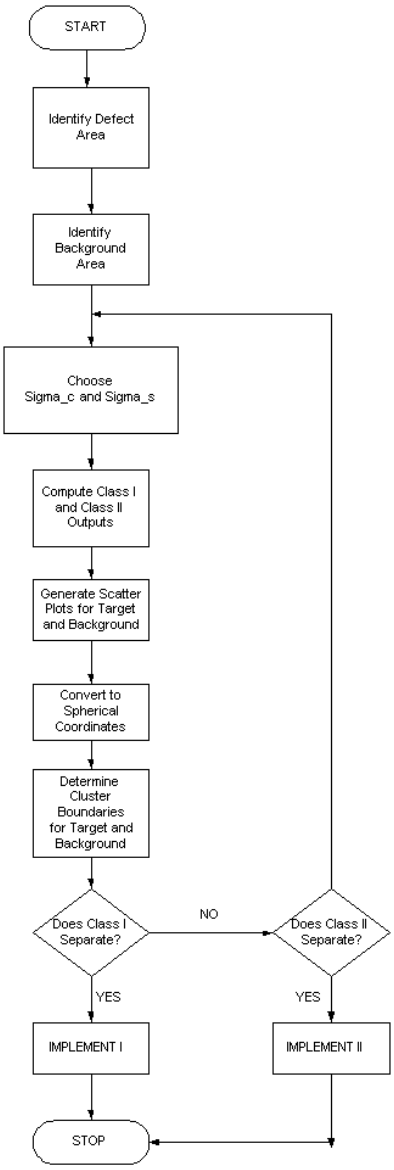

Figure 5.13: Flowchart illustrating the process of algorithm development

In this approach for example, the low pass version (output due to σ_s processing) of the image could be obtained in a variety of ways: an average of the image background could be used or the camera lens could be defocused thereby eliminating the need to conduct convolution operations to obtain these results. Additionally, the choice of σ_c could be chosen to filter noise or to reduce the effect of small regions that might not be significant. The system designer thus has the ability to modify the approach to obtain the results desired while being guided by the general principles outlined in the development of the approach. This would occur in the operations labelled IMPLEMENTATION I and IMPLEMENTATION II.

Another significant requirement in some applications, is the measurement of absolute color. The HVS, however, is considered a change detector and is not good at representing absolute image or scene parameters. The approach described has the same shortcoming, and, if this is a necessity would have to be addressed through the use of a reference standard where we would then be able to monitor the variation in the responses from the reference standard(s).

5.2.2 Space Transformation for Classification

In the process described in Chapter 3 and illustrated in Figure 3.1, we describe operations at Level 3 for making the final decisions for the problems being considered. At the rate of processing that is needed, it is considered here to be a simple classification operation.

Looking at the data for *Class I* we observe that this can be accomplished more easily by using a transformation that provides linear decision boundaries. This is achieved by the use of a spherical transformation of the processed space as given by

Equation (5.1). This is motivated by the fact that in the processed output space, the eigenvectors corresponding to the largest eigenvalues for the scatter data in the response space lie along a radial direction in the Cartesian representation. The data for Figure 5.8 in the transformed space is shown in Figure 5.14. It is observed that linear boundaries can be constructed for this data. This representation is used for conducting the classification hereafter. Similar observations are also made for the $Class\ II$ responses.

$$
\begin{aligned}
Radius &= \sqrt{R1^2 + R2^2 + R3^2} \\
Theta &= Tan^{-1}\left[\frac{abs(R2)}{abs(R1)}\right] \\
Phi &= Tan^{-1}\left[\frac{abs(R1)}{\sqrt{R1^2 + R2^2}}\right]
\end{aligned}
\tag{5.1}
$$

5.2.3 Example Applications

We will now give a brief description of the applications and the motivation for choosing them as demonstrations of the application of the approach. The first example is that of fan bone detection and has been discussed earlier in describing the development of the approach. The second example is one directed at the inspection of fruit. This is carried out at a rate an order of magnitude greater than the first example and also introduces concerns with the measurement of absolute color. The last example describes an application which is a combination of natural and manufactured products we are interested in the integrity of a package seal that could be contaminated with natural product such as juices and fats. Additionally this application is conducted

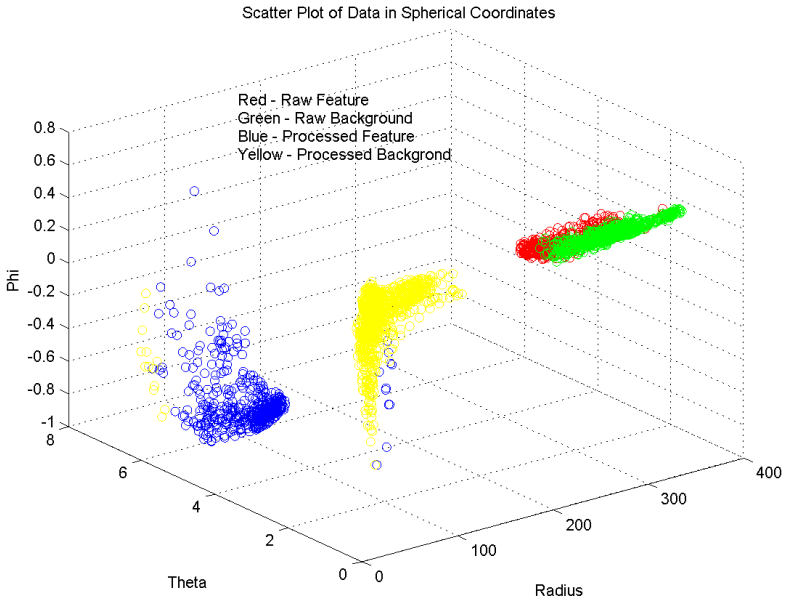

Figure 5.14: Scatter data from Figure 5.8 in spherical coordinates

Figure 5.15: Prototype fan bone detection system online at input to X-Ray imaging system

at fairly high resolution on the order of 1000 pixels per inch to enable us to locate defects of very small sizes on the order of 50 microns.

5.2.3.1 Sample Application 1 Fan bone Detection

The detection of surface fan bones is an integral part of the inspection process for deboned breast butterflies. These parts typically go by at rates of 30 to 60 pieces per minute. In this application the imaging cell acts as a front end for an x-ray system. The X-ray system has difficulty identifying fan bones as they are typically softer and thinner than other bones. This thus presents difficulties for the system to find the deeper more embedded bones as it is difficult to select X-ray energies to find all these bone types simultaneously. It was therefore decided to use visible imaging to find the fan bones as they typically occur on the surface. In Figure 5.15 we show a picture of the system on-line showing the imaging cell in the foreground with the X-ray imaging system in the background. The system consists of two 500 MHz PCs processing images from cameras monitoring each of the two lanes shown.

The process used on each PC is shown in Figure 5.16. Once the image is acquired color is analyzed using *Class I* processing; the goal being the initial identification of potential fan bone regions. In the next step, potential fan bone regions are filtered based on position and orientation. Because of the speed requirements, the algorithm was implemented by computing an average value for the image which was used as the surround and the original 'in focus' image as the center. Using the results of the initial classification, segmentation is conducted by the use of deformable contours (snakes) which use as the initial contour for segmentation the boundary of the binary regions formed as the result of the initial *Class I* processing. Once these regions are identified, then other features such as shape and color for the region are used as features in a second classification operation to improve our confidence in the final decision.

Sample outputs for this application for a typical butterfly are presented in Figure 5.17 and Figure 5.18 showing the results of the initial *Class I* classification which highlights the fan bone in red. This system operates with a detection accuracy of about 90 percent in this application across a variety of imaging configurations and was found to be more robust compared to using the raw image data. This is a significant improvement when compared to the 30 percent accuracy of the X-ray system alone. Most of the error in detection was attributed to part presentation, where the fan bones were either not visible (covered by meat or other material) or were in positions that presented small profiles to the imaging system so they did not match the size or shape of the typical fan bone.

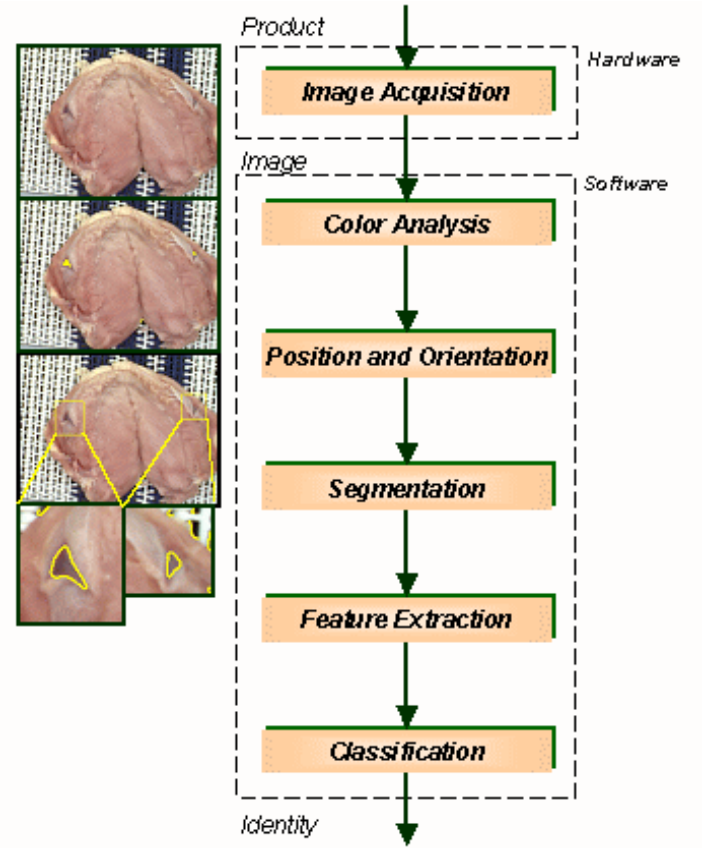

Figure 5.16: Flowchart for fanbone detection

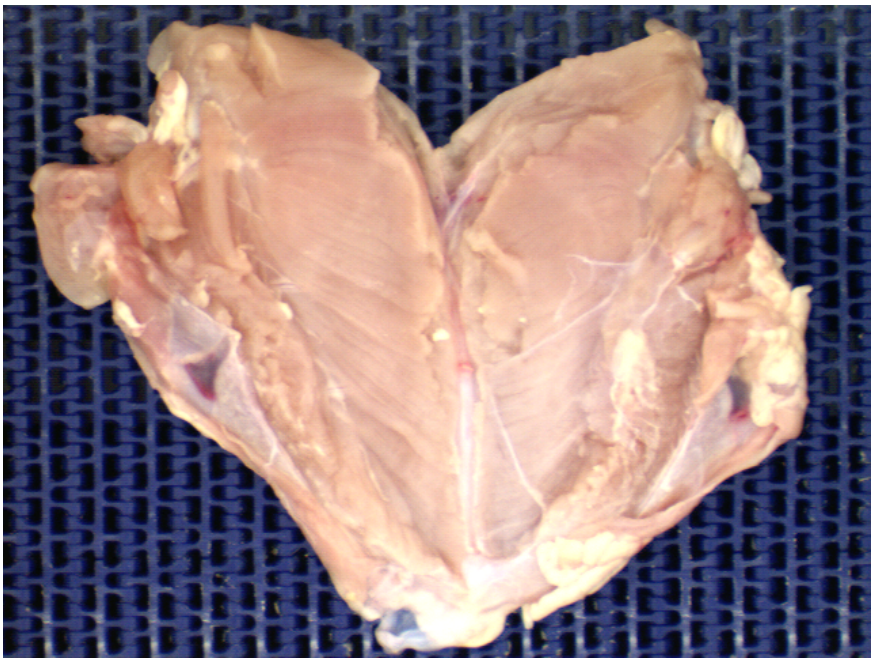

Figure 5.17: Sample fanbone image before processing

Figure 5.18: Sample fanbone image after processing

Figure 5.19: Picture showing lab prototype of the grapefruit inspection cell

5.2.3.2 Sample Application 2 - Grapefruit Inspection

Fruit inspection is a high speed labor intensive task that has to be conducted at speeds of up to 600 pieces of fruit per minute. This is another area where the variability of the product has made it difficult to implement automated solutions. A picture of the system used in the implementation is shown in Figure 5.19 and a block diagram of the system and its components shown in Figure 5.20 [38]. As shown, the system supports eight cameras looking at eight regions of the fruit. Four 500 MHz pentium computers process the image data with each computer supporting two cameras apiece. These machines (called clients) then provide summary data to the server computer that services the user interface and also the ESOP which controls the kickoff devices on the conveyor for sorting the defective fruit. A more detailed description of the system can be found in [38].

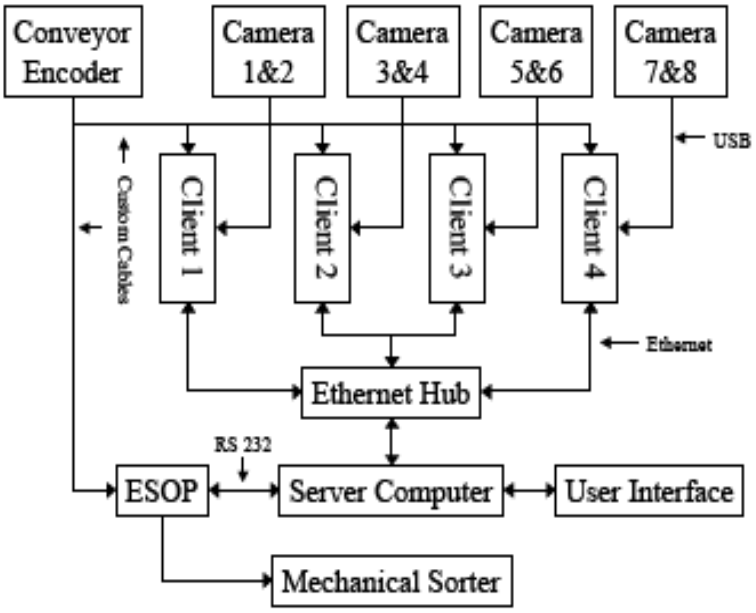

Figure 5.20: Block diagram of the grapefruit inspection cell illustrating its components

This application while similar to the fan bone problem in terms of being a natural product has several significant differences that demonstrates the robustness and wide applicability of the proposed approach. The differences include

- An artifact with a somewhat regular shape
- Overall different place in the color space
- Significantly higher speeds
- Different lighting scheme with multiple sources
- Spherical shape leading to non-uniform illumination

A sample defect that occurs on the grapefruit called a scar is shown in Figure 5.21 . The Class I and Class II scatter plots for the scar and background (normal

Figure 5.21: Sample grapefruit image with a scar defect

surface) are shown in Figure 5.22 and Figure 5.23 respectively. It is observed that we can fairly easily describe boundaries to separate the clusters.

The practical implementation of the solution required some modifications especially in light of the speed requirements. In a similar way to the fan bone problem and realizing that the surround is a low pass filtered version of the original image we decided to use reference balls as shown in Figure 5.24 to provide the surround response with the center response provided by the current image. The reference balls are generated by having them painted to match the color of acceptable fruit; this way, the reference balls provide a color standard as is required in some implementations and additionally serves as a template for quantifying shape deviations. Sample images and the output images in which we detect scarring of the fruit is shown in Figure 5.25 where the fruit to be graded are shown on the right and the scar areas are identified in the areas in blue in the images on the left. Tests showed that this

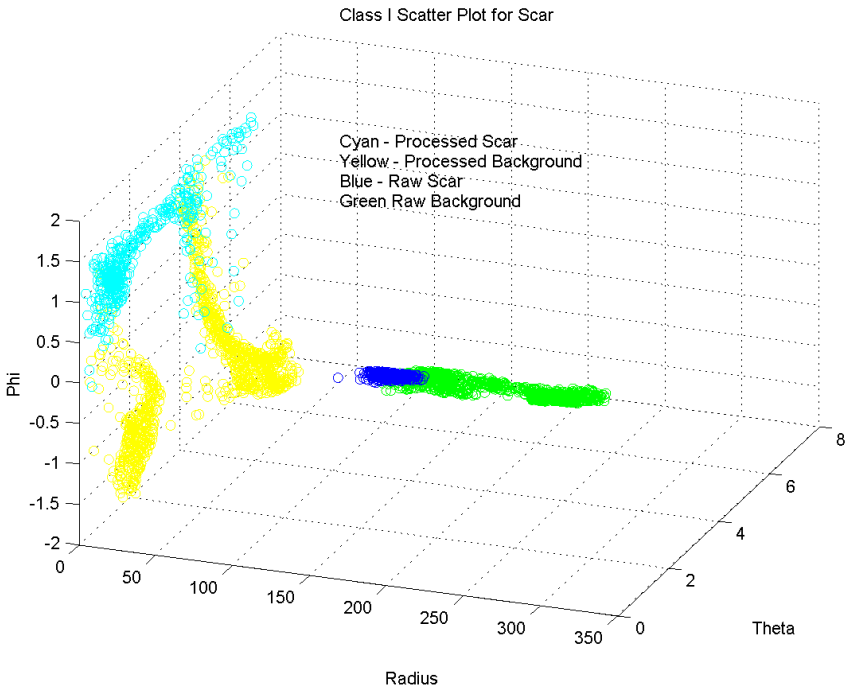

Figure 5.22: Scatter plots for Class I outputs for the scar sample image

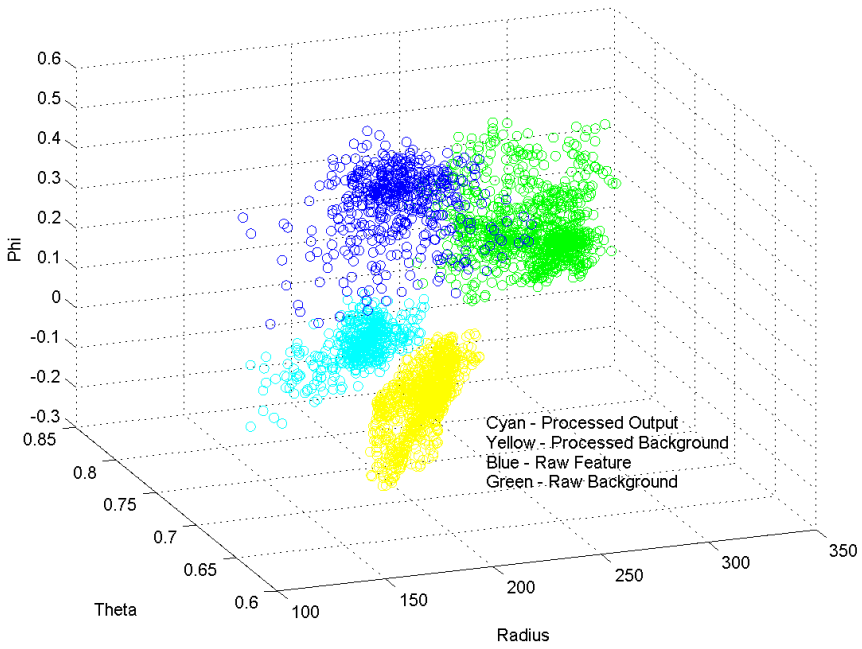

Figure 5.23: Scatter plots for Class II outputs of grapefruit scar data

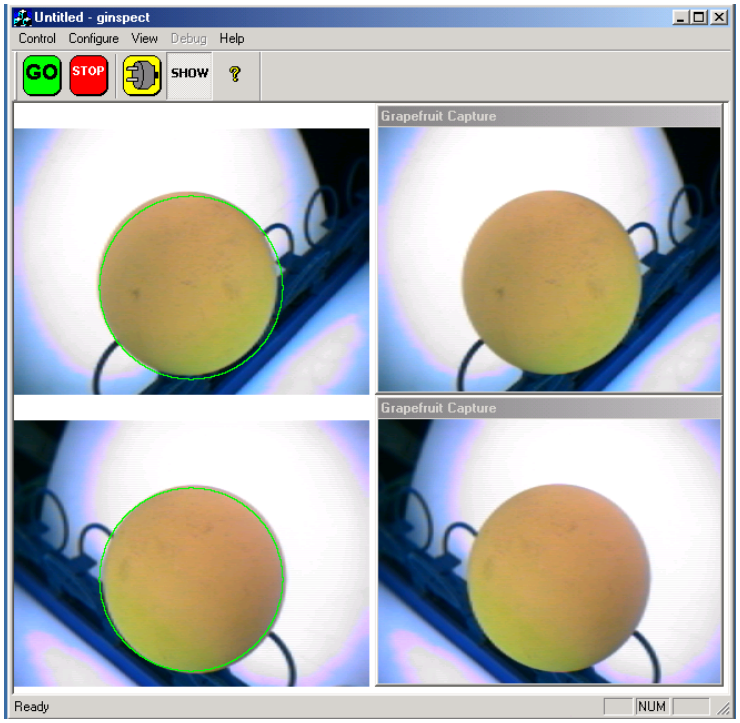

Figure 5.24: Reference balls used to provide the surround response

system matched the performance of human graders at the desired rates in laboratory tests. In addition, this approach also demonstrated the ability to function under a variety of lighting configurations without changes in the algorithms.

5.2.3.3 Sample Application 3 Package Inspection

Package inspection of seals have taken on added importance over the past few years. Concerns about the integrity of package seals are important mainly for food safety considerations as it is the main protection mechanism for the product once it leaves the producer and is sent to the consumer. This process is currently done manually

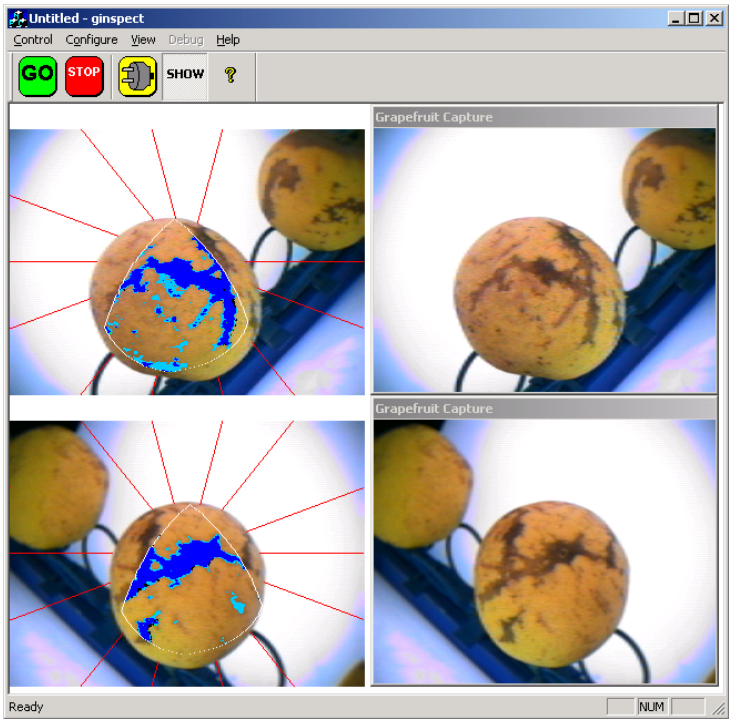

Figure 5.25: Sample input and output images for detecting scar on the surface of a grapefruit

and there is strong motivation to automate this inspection as it enhances the possibilities to automate downstream processes making the whole operation safer and more reliable. This application is a little different than the two described previously in that it is:

- Done at a moderate rate of speed (30 to 60 packages per minute)
- Combination of natural and manufactured components (food, plastic,foam)
- High resolution imaging to find small defects (on order of 50 microns)
- Occurs at a different place in the color space

A sample seal with a defect is shown in Figure 5.26. The seal is made by using heat and pressure to bond plastic film to an underlying foam tray. This image is a high resolution picture of the seal of what is called a lidded MAP (Modified Atmosphere Package). Defects can occur either due to the formation of a defective seal due to a malfunction of the machinery or contamination of the seal through the inadvertent deposition of material in the seal area. The scatter plots for normal and contaminated areas of the seal are presented in Figure 5.27 and Figure 5.28 for packages with a white and black tray respectively.

A sample application developed using the approach described above is shown in Figure 5.29. In a similar vein to the previous example and to reduce processing time the surround processing is executed on a good seal and saved. The current image is then used as the center response with $Class\ I$ classification conducted as before. It can be seen that we are able to identify the contamination (the reddish area on the right) as shown in the output image (white area on the right).

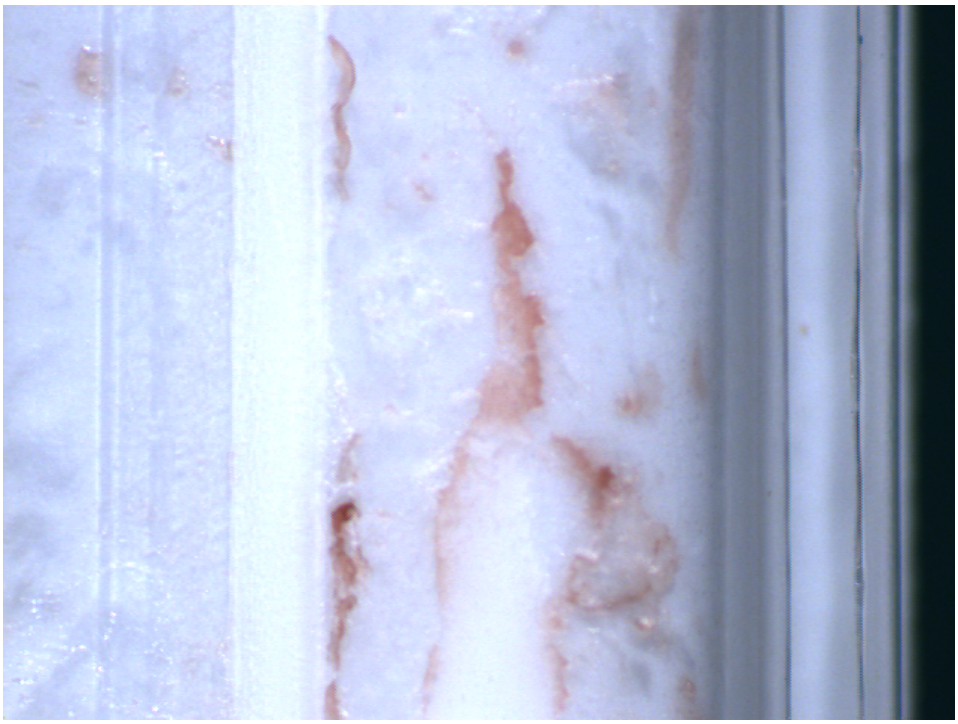

Figure 5.26: Package seal with contamination

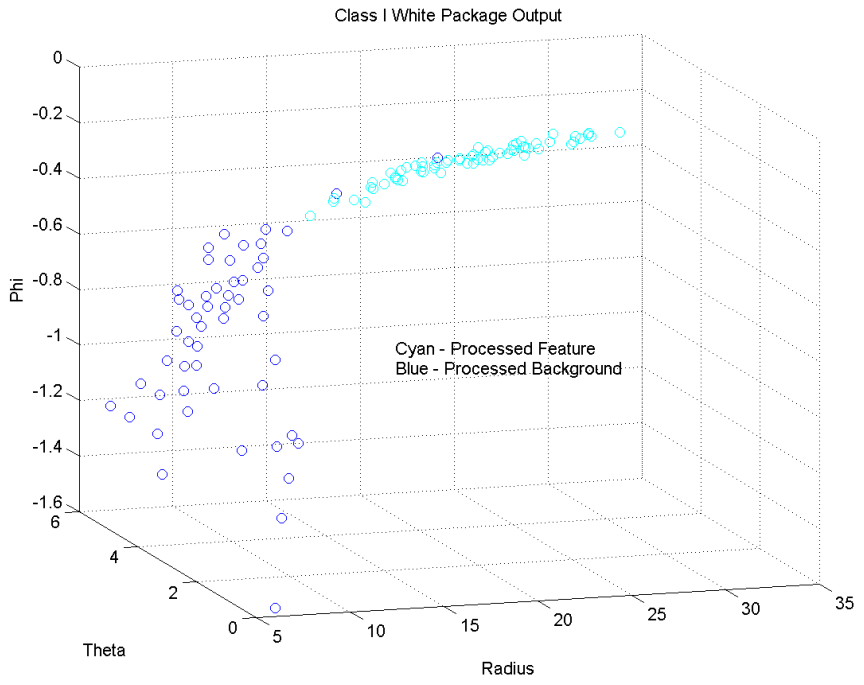

Figure 5.27: Scatter plots for Class I ouputs for defective seal in a white tray

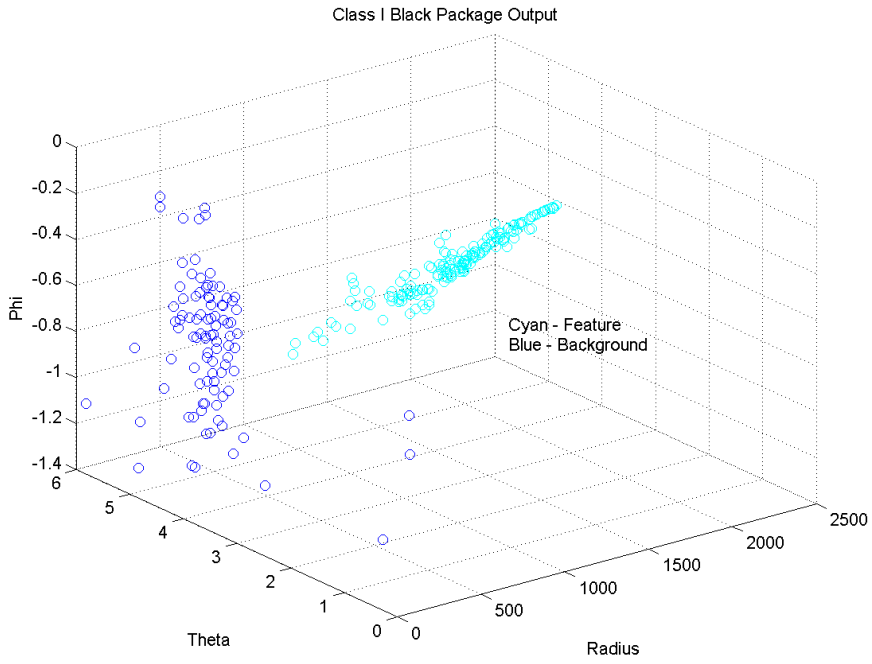

Figure 5.28: Scatter plots for Class I output for a defective seal on a black tray

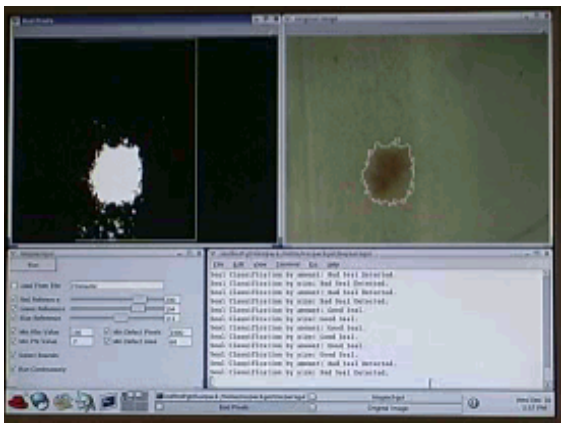

Figure 5.29: Detection of a contaminated seam on lidded map package

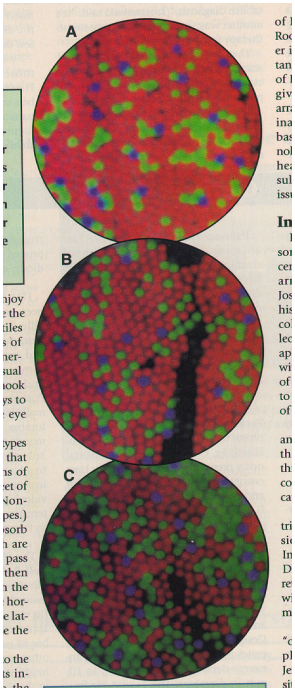

Figure 5.30: Distribution of S, M and L wavelength sensors in the eye of three different people with normal color vision [39]

5.3 Effect of visual deficiencies

Humans do not see color in the same way, as we all have somewhat different responses to the sensory inputs. The images in Figure 5.30 [39] shows the distribution of LMS sensors in three people that were tested to have normal color vision. Additionally some can have significant defects. The image in Figure 5.31 displays how a Deutan (someone with a poor response in the M cones) might see the prototype fan image in Figure 5.1(b). The responses for this case are shown in Table 5.11 and it is seen that discrimination is still possible.

Figure 5.31: Fan bone prototype image as would be seen by a Deutan

Table 5.11: Response outputs for Deutan

Fan Prototype	Deutan	
	Feature	Background
R1	-5.136	10.82
R2	0	0
R3	1.823	4.695

5.4 Summary

We have proposed an approach based on models that describe some aspects of the functioning of the human visual system. From these models we have deduced features to be extracted from images that are useful for identifying defects in natural products We have demonstrated the application of the technique on three different problems with no significant changes in the approach. The resulting solutions have been shown to be robust to the expected changes in the environment and the variability of the product. The approach using the Class I computations consists of looking at the difference in color space between a low pass and high pass version of the image and carrying out classification in this representation. The Class II representation seems to allow for better color discrimination but, at higher computational cost.

Chapter 6

Conclusions and Recommendations

6.1 Conclusions

The thrust of this thesis is to examine approaches using human visual models to develop effective machine vision solutions. Specifically, we set out in this effort to develop a method that could provide a technique to assist in the development of machine vision algorithms for defect detection especially for natural products. The specifications for many problems are somewhat subjective in nature and require some training of inspectors. Humans are currently the most effective means of addressing these quality control and inspection tasks. They therefore seemed a natural model to use for insights into the development of machine vision solutions to address these kinds of problems. It has been demonstrated, however, that the approach is also extensible to other products with significant natural variations as occurs for example in package seal inspection.

Specific findings of this thesis are summarized below:

- We have shown that starting with the basic descriptions and results from the areas of biology, physiology and human vision research we have been able to identify operations on images that are able to extract features that are useful

for classifying image regions. A major element, is the encoding of contrast and the mechanism for its computation through the concept of a receptive field. These results give performance that is robust under many of the changes in conditions that would occur under the typical industrial installation. We have extended the earlier research conducted, and models developed mostly in the analysis of monochromatic imagery to the analysis of color imagery. This was accomplished by using the concept of the receptive field and its representation as a difference of Gaussians.

- We have developed a systematic approach towards the choice of system parameters that involves posing the problem as one of optimization through the use of mathematical models describing the response of the receptive fields. This is accomplished by choosing parameters to maximize the distances between clusters of interest in the response space.

- We have defined two combinations of center and surround responses called *Class I* and *Class II*, which were obtained by defining the center surround using the trichromatic and opponent color theories. The Class I response provides an efficient way to encode edge information and to categorize larger color variances. The *Class II* responses, on the other hand, offer a means to enhance smaller color differences.

- We have demonstrated that the transformation of the data from the *Class I* and *Class II* response space using a spherical transformation produced a representation that allowed for the use of linear decision boundaries for classification. This helps to simplify the development and implementation.

- We have demonstrated the applicability of the approaches to three application problems, meat grading, fruit sorting, and package inspection. Through these application examples we showed how the technique can be applied to new problems by acquiring example images that are able to show the difference between the defective and normal product, and then to design decision boundaries based on how the defects clustered in the response space.

- Finally, we also have laid a foundation on which to continue to build as we learn and understand more about the functions of the human visual system. We anticipate that as we understand more about these other processes, particularly at the higher levels of the brain, a similar approach could be used to exploit this new knowledge to develop more sophisticated algorithms to solve machine vision problems.

In summary, we have developed a systematic approach to guide the process of algorithm development drawing from the knowledge of the biological principles that govern the operation of the human visual system. This approach is especially applicable in domains where humans are currently the sensing modes of choice for example in the inspection and sorting of natural products, but is also shown to be useful in other applications.

6.2 Future Work

There are many potential avenues to pursue in terms of other mechanisms that could be useful in guiding the development of algorithms. In particular we have looked

here mostly at the low level operations and have not ventured into some of the higher level functions of the brain This is definitely a fertile area to pursue and could produce the more significant results in the long run; there is still much work to be done in exploiting the lower level functions however. The approach described here for the classification of regions could be extended for example to the development of general color image classification algorithms. This could be a useful approach to the segmentation of other natural scenes such as might be needed for robotic guidance in the detection of fruit for picking or automated vehicle control. It might also be possible to extend the approach to the analysis of hyperspectral imagery. Exploration of just noticeable contrast thresholds in more complicated scenes that could provide useful results for noise filtering.

It is also known that there is a temporal aspect to the receptive field responses; the impact of this behavior and the potential benefits of these extensions should be explored to determine their usefulness.

The surround responses for the receptive fields was obtained from two of the three main theories that describe color vision today, the trichromatic theory and the opponent theory. The opponent process identified earlier is also thought to facilitate the recovery of spectra. The basic question would then be, can spectra be recovered and would multispectral analysis enhance our ability to identify defects? It is currently believed that this is one of the higher level brain functions and helps to explain the phenomenon of color constancy; the idea being that we are able to encode the reflectance properties of the object as opposed to the illumination and that we are able to extract this information under different illumination schemes.

Stereo vision and correspondence is also a potential area for investigation. Looking

for example, at the receptive fields that respond to stereo could provide useful results to assist in the determination of correspondence in stereo images and would address many problems related to 3D sensing.

The development of imaging sensors with some of the low level color processing algorithms integrated could also aid in lowering the costs related to implementing many of these systems.

Appendix A

Calculation of Responses for Sample Images

Starting with Equation (4.18) and Equation (4.19) they can be rewritten as shown in Equation (A.1) and Equation (A.2) with the substitutions below.

$$C = 4\gamma q\left(1 + 2\varepsilon_{2c} + \varepsilon_{1c}\right) + 4\gamma_{2c} p\left(3\varepsilon_{1c} + 2\varepsilon_{2c}\right) \tag{A.1}$$

$$S = 4\gamma_{2s}\varepsilon_{1s} q + 4\left(\gamma_{2s} + 3\gamma_{2s}\varepsilon_{1s} + 4\gamma_{2s}\varepsilon_{2s}\right) p \tag{A.2}$$

$$+8\left(\gamma_{2s}\varepsilon_{2s} + \gamma_{2s}\varepsilon_{1s}\right) q + 8(\gamma_{2s} + 3\gamma_{2s}\varepsilon_{2s} + 3\gamma_{2s}\varepsilon_{1s})p$$

$$+4\gamma q\left(1 + 2\varepsilon_{2c} + \varepsilon_{1c}\right) + 4\gamma_{2c} p\left(3\varepsilon_{1c} + 2\varepsilon_{2c}\right) p \tag{A.3}$$

where:

$\gamma_{2c} = \frac{1}{2\pi\sigma_c^2}$, $\varepsilon_{2c} = \exp(-\frac{1}{2\sigma^2})$, $\varepsilon_{1c} = \exp(-\frac{1}{\sigma^2})$, $\gamma_{2s} = \frac{1}{2\pi\sigma_s^2}$ $\varepsilon_{2s} = \exp(-\frac{1}{2\sigma_s^2})$, $\varepsilon_{1s} = \exp(-\frac{1}{\sigma_s^2})$

For a given image Substitute and set $q = (1 + r)\, p$, J can be rewritten by substituting C and S from Equation (A.1) and Equation (A.2) into Equation (4.15) to obtain the result in Equation (A.4).

$$J = 4\gamma(1+r)p\,(1+2\varepsilon_{2c}+\varepsilon_{1c}) + 4\gamma_{2c}p\,(3\varepsilon_{1c}+2\varepsilon_{2c}) - (4\gamma_{2s}\varepsilon_{1s}(1+r)p$$
$$+ 4\,(\gamma_{2s}+3\gamma_{2s}\varepsilon_{1s}+4\gamma_{2s}\varepsilon_{2s})\,p + 8\,(\gamma_{2s}\varepsilon_{2s}+\gamma_{2s}\varepsilon_{1s})\,(1+r)p$$
$$+ 8(\gamma_{2s}+3\gamma_{2s}\varepsilon_{2s}+3\gamma_{2s}\varepsilon_{1s})p + 4\gamma_{2s}(1+r)p\,(1+2\varepsilon_{2c}+\varepsilon_{1c})$$
$$+ 4\gamma_{2s}p\,(3\varepsilon_{1c}+2\varepsilon_{2c})\,p) \quad \text{(A.4)}$$

We will identify the three Class I response types as $J_1,\ J_2$, and J_3 corresponding to the RGB or SML sensor types. The choices for σ_c and σ_s will determine the magnitude of the output response.

We can write the color response function for J_4 by modifying Equation (A.4) to get Equation (A.5).

$$J_4 = 4\gamma_{2c}q_r\,(1+2\varepsilon_{2c}+\varepsilon_{1c}) + 4\gamma_{2c}p_r\,(3\varepsilon_{1c}+2\varepsilon_{2c}) - (4\gamma_{2s}\varepsilon_{1s}(q_r-q_g)$$
$$+ (4\gamma_{2s}+3\gamma_{2s}\varepsilon_{1s}+4\gamma_{2s}\varepsilon_{2s})(p_r-p_g) + 8\,(\gamma_{2s}\varepsilon_{2s}+\gamma_{2s}\varepsilon_{1s})\,(q_r-q_g)$$
$$+ 8(\gamma_{2s}+3\gamma_{2s}\varepsilon_{2s}+3\gamma_{2s}\varepsilon_{1s})(p_r-p_g) + 4\gamma_{2s}(q_r-q_g)\,(1+2\varepsilon_{2s}+\varepsilon_{1s})$$
$$+ 4\gamma_{2s}(p_r-p_g)\,(3\varepsilon_{1s}+2\varepsilon_{2s})) \quad \text{(A.5)}$$

Similarly we can obtain J_5 and J_6 as shown in Equation (A.6) and Equation (A.7) respectively. These are the forms used for computing these responses.

$$J_5 = 4\gamma_{2c}q_g\left(1+2\varepsilon_{2c}+\varepsilon_{1c}\right)+4\gamma_{2c}p_g\left(3\varepsilon_{1c}+2\varepsilon_{2c}\right)-\left(4\gamma_{2s}\varepsilon_{1s}(q_r-q_g)\right.$$
$$+\left(4\gamma_{2s}+3\gamma_{2s}\varepsilon_{1s}+4\gamma_{2s}\varepsilon_{2s}\right)(p_r-p_g)+8\left(\gamma_{2s}\varepsilon_{2s}+\gamma_{2s}\varepsilon_{1s}\right)(q_r-q_g)$$
$$+8(\gamma_{2s}+3\gamma_{2s}\varepsilon_{2s}+3\gamma_{2s}\varepsilon_{1s})(p_r-p_g)+4\gamma_{2s}(q_r-q_g)\left(1+2\varepsilon_{2c}+\varepsilon_{1c}\right)$$
$$\left.+4\gamma_{2c}(p_r-p_g)\left(3\varepsilon_{1c}+2\varepsilon_{2c}\right)\right) \quad \text{(A.6)}$$

$$J_6 = 4\gamma_{2c}q_b\left(1+2\varepsilon_{2c}+\varepsilon_{1c}\right)+4\gamma_{2c}p_b\left(3\varepsilon_{1c}+2\varepsilon_{2c}\right)-\left(4\gamma_{2s}\varepsilon_{1s}(q_b-q_r+q_g)\right.$$
$$+\left(4\gamma_{2s}+3\gamma_{2s}\varepsilon_{1s}+4\gamma_{2s}\varepsilon_{2s}\right)(p_b-p_r+p_g)+8\left(\gamma_{2s}\varepsilon_{2s}+\gamma_{2s}\varepsilon_{1s}\right)(q_b-q_r+q_g)$$
$$+8(\gamma_{2s}+3\gamma_{2s}\varepsilon_{2s}+3\gamma_{2s}\varepsilon_{1s})(p_b-p_r+p_g)+4\gamma_{2s}(q_b-q_r+q_g)\left(1+2\varepsilon_{2s}+\varepsilon_{1s}\right)$$
$$\left.+4\gamma_{2s}(p_b-p_r+p_g)\left(3\varepsilon_{1s}+2\varepsilon_{2s}\right)\right) \quad \text{(A.7)}$$

Set

$$q_r = p_r + r_1 p_r \quad \text{(A.8)}$$

$$q_g = p_r + r_2 p_r \quad \text{(A.9)}$$

$$p_g = p_r + r_3 p_r \quad \text{(A.10)}$$

For J_5 define

$$p_r = p_g + r_4 p_g \tag{A.11}$$

$$q_g = p_g + r_5 p_g \tag{A.12}$$

$$q_r = p_g + r_6 p_g \tag{A.13}$$

Substituting in Equation (A.6) we get

$$\begin{aligned} J_5 = 4\gamma_{2c}(1+r_5)p_g\left(1+2\varepsilon_{2c}+\varepsilon_{1c}\right) + 4\gamma_{2c}p_g\left(3\varepsilon_{1c}+2\varepsilon_{2c}\right) - \left(4\gamma_{2s}\varepsilon_{1s}(r_6-r_5)\right. \\ + (4\gamma_{2s}+3\gamma_{2s}\varepsilon_{1s}+4\gamma_{2s}\varepsilon_{2s})(r_4 p_g) + 8\left(\gamma_{2s}\varepsilon_{2s}+\gamma_{2s}\varepsilon_{1s}\right)(r_6-r_5) \\ + 8(\gamma_{2s}+3\gamma_{2s}\varepsilon_{2s}+3\gamma_{2s}\varepsilon_{1s})(r_4 p_g) + 4\gamma_{2s}(r_6-r_5)\left(1+2\varepsilon_{2c}+\varepsilon_{1c}\right) \\ \left. + 4\gamma_{2c}(r_4 p_g)\left(3\varepsilon_{1c}+2\varepsilon_{2c}\right)\right) \end{aligned} \tag{A.14}$$

Similarly for J_6 define

$$p_r = p_b + r_7 p_b \tag{A.15}$$

$$q_b = p_b + r_8 p_b \tag{A.16}$$

$$p_g = p_b + r_9 p_b \tag{A.17}$$

$$q_r = p_b + r_{10}p_b \tag{A.18}$$

$$q_g = p_b + r_{11}p_b \tag{A.19}$$

Substituting in Equation (A.7) we get

$$\begin{aligned} J_6 = (&4\gamma_{2c}(1+r_8)\,(1+2\varepsilon_{2c}+\varepsilon_{1c}) + 4\gamma_{2c}\,(3\varepsilon_{1c}+2\varepsilon_{2c}) - (4\gamma_{2s}\varepsilon_{1s}(1+r_8-r_{10}+r_{11}) \\ &+ (4\gamma_{2s}+3\gamma_{2s}\varepsilon_{1s}+4\gamma_{2s}\varepsilon_{2s})(1+r_9-r_7) + 8\,(\gamma_{2s}\varepsilon_{2s}+\gamma_{2s}\varepsilon_{1s})\,(1+r_8-r_{10}+r_{11}) \\ &+ 8(\gamma_{2s}+3\gamma_{2s}\varepsilon_{2s}+3\gamma_{2s}\varepsilon_{1s})(1+r_9-r_7) + 4\gamma_{2s}(1+r_8-r_{10}+r_{11})\,(1+2\varepsilon_{2c}+\varepsilon_{1c}) \\ &+ 4\gamma_{2c}(1+r_9-r_7)p_b\,(3\varepsilon_{1c}+2\varepsilon_{2c})))p_b \end{aligned} \tag{A.20}$$

Appendix B

Camera Model Equations for Gains

$$
\begin{aligned}
V_s &= G_{c1}e \\
V_{wb} &= G_{c1}G_2e && \text{(B.1)} \\
V_g &= (G_{c1}G_{c2})^{\gamma_c}e^{\gamma_c} && \text{(B.2)} \\
V_o &= G_{c3}(G_{c1}G_{c2})^{\gamma_c}e^{\gamma_c} && \text{(B.3)} \\
V_o &= G_{cT}e^{\gamma_c} && \text{(B.4)} \\
\log(V_o) &= \log(G_{cT}e^{\gamma_c}) && \text{(B.5)} \\
\log(V_o) &= \log(G_{cT}) + \log(e^{\gamma_c}) && \text{(B.6)} \\
\log(V_o) &= \log(G_{cT}) + \gamma_c \log(e) && \text{(B.7)}
\end{aligned}
$$

We now determine the parameters of the above model using experimental data to do a least squares curve fit. With this model we are able to predict grey scale output for particular input energies to the senor.

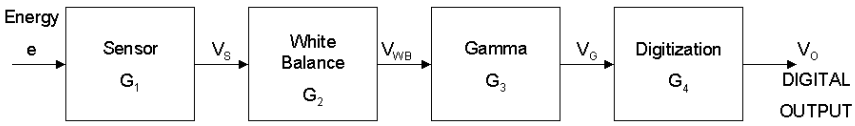

Figure B.1: Block diagram illustrating the imaging process

Appendix C

Camera Model

Light energy from the external world is collected by an optical system and then imaged on a sensor. The energy absorbed at each point is then used to generate a digitized representation of the scene imaged on the sensor.

The formulation presented is derived from [40]. The geometry for using the thin lens imaging formulation is shown in Figure C.1. The apparent area of the image patch as seen from the center of the lens is given by $dI_m \cos\alpha$.

$$Solid\ Angle I_m = \frac{dI_m \cos\alpha}{r^2} \tag{C.1}$$

$$\cos\alpha = \frac{f}{r_I} \tag{C.2}$$

$$r_I = \frac{f}{\cos\alpha} \tag{C.3}$$

Substitute for r_I and we get

$$Solid\ Angle I_m = \frac{dI_m \cos^3\alpha}{f^2} \tag{C.4}$$

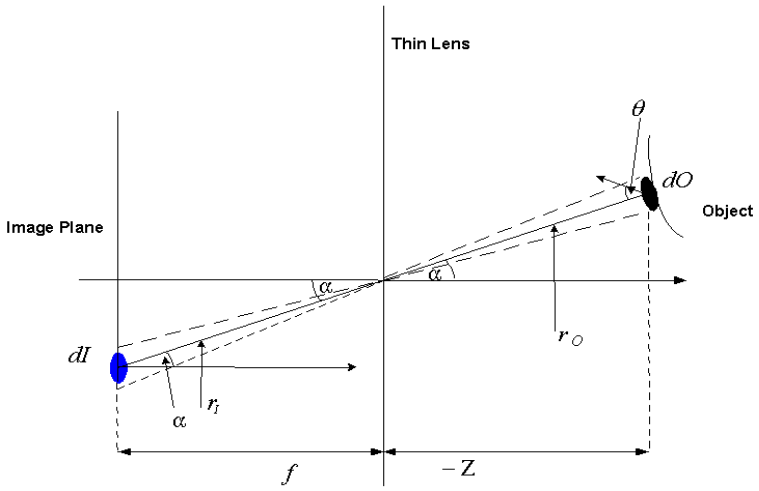

Figure C.1: Imaging geometry using a thin lens

Similarly for the object patch we get

$$Solid\ AngleO = \frac{dO\cos^3\theta}{r_o^2} \tag{C.5}$$

$$r_O = \frac{z}{\cos\alpha} \tag{C.6}$$

$$Solid\ AngleO = \frac{dO\cos\theta\cos^2\alpha}{z^2} \tag{C.7}$$

$$Solid\ AngleO = Solid\ AngleI_m \tag{C.8}$$

$$\frac{dI_m\cos^3\alpha}{f^2} = \frac{dO\cos\theta\cos^2\alpha}{z^2} \tag{C.9}$$

Rearranging we get

$$\frac{dO}{dI_m} = \frac{z^2\cos\alpha}{f^2\cos\theta} \tag{C.10}$$

The amount of light that passes through the lens and reaches the sensor is determined by the $F - number$ of the lens, therefore if the diameter of the lens is d then its area is:

$$A_L = \frac{\pi d^2}{4} \tag{C.11}$$

This implies that the projected area to the object patch is

$$A_P = \frac{\pi d^{\,2} \cos\alpha}{4} \tag{C.12}$$

The solid angle subtended by the lens on the patch is therefore

$$\frac{\pi d^2 \cos\alpha}{4r_0^2} = \frac{\pi d^2 \cos^3\alpha}{4z^2} = \Omega \tag{C.13}$$

Power of light from the object patch passing through the lens is

$$P = LdO\Omega\cos\theta \tag{C.14}$$

$$where\ L_q\ is\ the\ radiance\ of\ the\ surface \tag{C.15}$$

This results in the irradiance of the image patch Φ as shown below

$$\Phi = \frac{dP}{dI_m} \tag{C.16}$$

This implies that

$$\Phi = \frac{L_q dO\pi d^2 \cos^3\alpha\cos\theta}{dI4z^2} \tag{C.17}$$

Substituting for $\frac{dO}{dI}$ in Equation (C.10)

$$\Phi = \frac{L_q \pi d^2 \cos^4 \alpha}{4f^2} \tag{C.18}$$

With the $F - number$ defined as $\frac{d}{f}$ we get

$$\Phi = \frac{L_q \pi \cos^4 \alpha}{4F^2} \tag{C.19}$$

But $L_q = L_q(\lambda)$ and Energy E

$$E = \Phi dt dA \tag{C.20}$$

This implies that the energy for a sensor with sensitivity $S_i(\lambda)$ we get

$$E_i = \frac{\pi \cos^4 \alpha A_d t_{int}}{4F^2} \int_{\lambda_1}^{\lambda_2} S_i(\lambda) L_q(\lambda) d\lambda \tag{C.21}$$

The basic equation that governs the response of a camera is shown in Equation C.21 we then adapt this to reflect the three sensors in a color camera; this is then used to simulate the responses for cameras with the simulation tool described in Appendix D.

Appendix D

Simulation Tool

The main screen for the program is shown in Figure D.1 and is used to evaluate and compare the resulting output representations by modeling the sensor response in the case of a camera and the retinal response in the case of humans. It also makes it possible then to also evaluate the effect of deficiencies in the eye and how this might affect the resulting images. The program allows you to choose the system that you want to evaluate and then to choose a spectral input for that system. The result is a color display that shows how that color would be perceived by these different systems. These outputs of themselves do not tell us how images of real scenes will be processed and interpreted it only tells us the degree of match that would occur in the representation from each system.In order that we can reasonably develop algorithms models of the camera system and the human visual system were developed to evaluate the differences to be expected in image formation and their effect on the performance of algorithms. What kinds of contrasts should be expected. In particular for the same spectra are their significant difference in the features? Or, is there a set of invariant features that could be utilized. Two systems will be compared in terms of their representation at Level I. These will be the human system and artificial systems, in this case cameras. To facilitate these experiments a program was written to generate output representations under different conditions. The user interface is

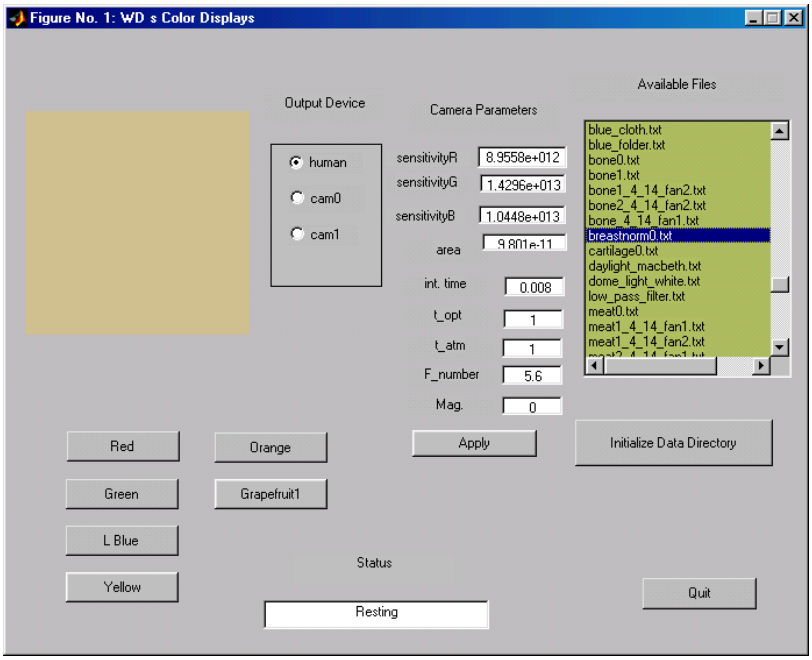

Figure D.1: Program to generate values for artificial images

shown in Figure D.1. We will first look at data for a camera. The inputs to the model are spectroradiometric data representing areas on interest in a scene. Sample spectroradiometric data is shown in Figure D.2 and sample camera response curves in Figure D.3.

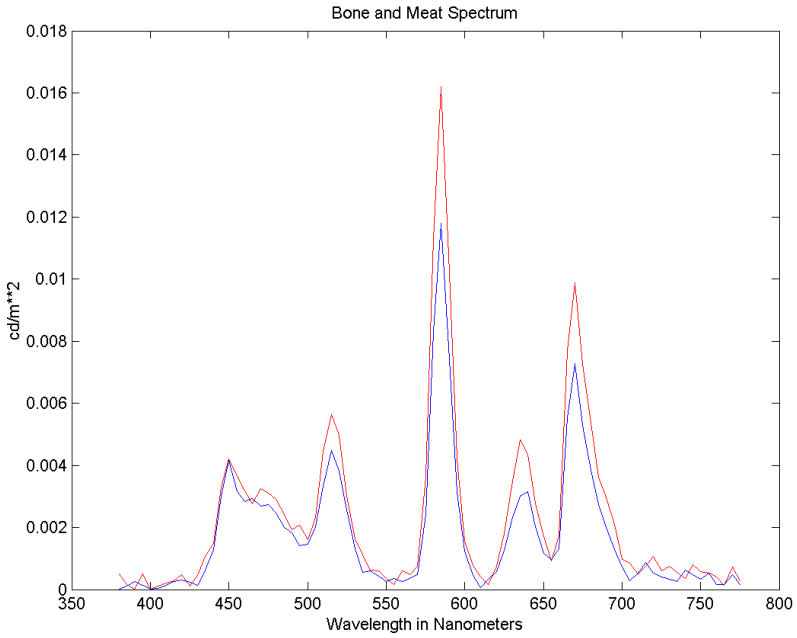

Figure D.2: Spectra of light reflected from meat (red) and bone (blue)

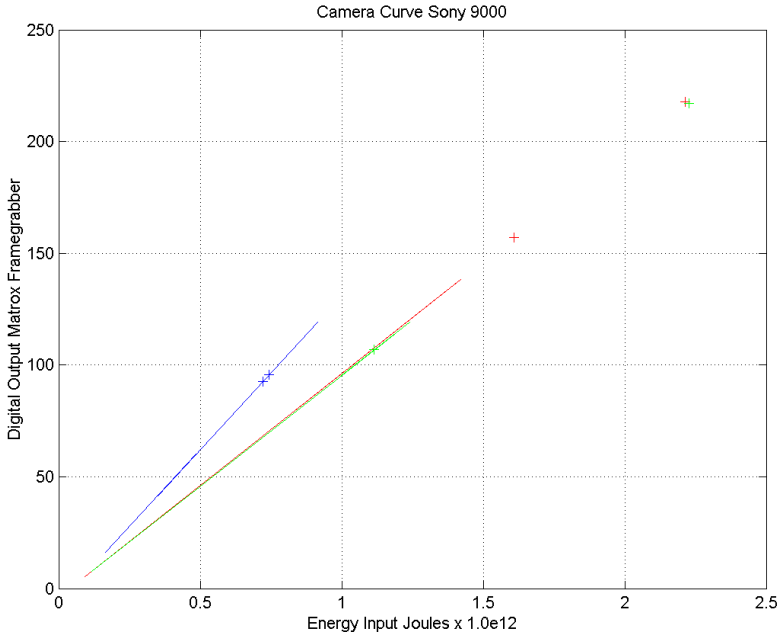

Figure D.3: Transformation Curves for the Sony 9000 Camera

Appendix E

General Formula for Gaussian Kernel

We would also like to have a general description for $\boldsymbol{\beta}$ to accomodate filters of different sizes, for example with a $5x5$ filter we would

have

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_8 & \beta_5 & \beta_4 & \beta_5 & \beta_8 \\ \beta_5 & \beta_2 & \beta_1 & \beta_2 & \beta_5 \\ \beta_4 & \beta_1 & \beta_0 & \beta_1 & \beta_4 \\ \beta_5 & \beta_2 & \beta_1 & \beta_2 & \beta_5 \\ \beta_8 & \beta_5 & \beta_4 & \beta_5 & \beta_8 \end{bmatrix} \tag{E.1}$$

Where

$$\beta_i(\sigma) = \frac{1}{2\pi\sigma^2} \exp(-\frac{1}{2}(\frac{i}{\sigma^2})) \tag{E.2}$$

For a general kernel with center pixel (c_1, c_2) and index $(1, 1)$ in top left, then the values for $\boldsymbol{\beta}$ are given by

$$\beta(i, j) = \beta(\sigma)_{(i-c_1)^2+(j-c_2)^2} \tag{E.3}$$

with the use of Equation (E.2) the values for these kernels can be computed.

Bibliography

[1] Mark Graves and Bruce Batcherlor. *Machine Vision for the Inspection of Natural Products*. Springer-Verlag, London, 2004.

[2] Tincher W.C., Daley W. D., and Holcombe W. Detection and removal of fabric defects. *International Journal of Clothing Science and Technology*, 4(2/3):54–65, 1992.

[3] Nello Zuech. *Applying Machine Vision*. John Wiley and Sons, New York, 1988.

[4] David W. Arathorn. *Map-Seeking Circuits in Visual Cognition*. Stanford University Press, Stanford, California, 2002.

[5] Rhodes M. Computers in surgery and therapeutic procedures. *IEEE Computer*, pages 20–23, January 1996.

[6] Yuzheng W., Doi Kunio, Giger Maryellen L., and Nishikawa Robert M. Computerized detection of clustered microcalcificatios in digital mammograms: Applications of artificial neural networks. *Med. Phys*, 19(3):555–560, May/June 1992.

[7] Marr David. *Vision, A Computational Investigation Into the Human Representation and Processing of Visual Information*. W. H. Freeman and Company, New York, 1982.

[8] Grinsom W. E. L. *From Images to Surfaces A Computational Study of the Human Early Visual System*. MIT Press, Boston, MA, 1981.

[9] Rogers E. *Visual Interaction: A Link Between Perception and Problem-Solving*. PhD thesis, School of Information and Computer Science, Georgia Institute of Technology, November 1992.

[10] Doll T. J., McWhorter S. W., Wasilewski A. A., and Schmieder D. E. Georgia tech vision (GTV) model version GTV96 analysís manual. 1997. Prepared under GTRI contract DAAJ02-92-C-0044.

[11] Gershon R. *The Use of Color in Computational Vision*. PhD thesis, , Department of Computer Science, University of Toronto, 1987.

[12] Belkacem-Boussaid K. and Beghdadi A. A new image smoothing method based on a simple model of spatial processing in the early stages of human vision. *IEEE Transactions on Image Processing*, 9(2):220–226, February 2000.

[13] Samir Shah and Martin D. Levine. Visual information processing in primate cone pathways-part i: A model. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, 26(2):259–274, April 1996.

[14] Samir Shah and Martin D. Levine. Visual information processing in primate cone pathways-part II: Experiments. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, 26(2):275–289, April 1996.

[15] T.V. Papathomas, R. S. Kashi, and A. Gorea. A human based computational model for chromatic texture segregation. *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 27(3):428–440, June 1997.

[16] Claudio M. Privitera and Lawrence W. Stark. Human-vision-based selection of image processing algorithms for planetary exploration. *IEEE Transactions on Image Processing*, 12(8):917–923, August 2003.

[17] Ko Sakai and Leif H. Finkel. A shape-from-texture algorithm based on human visual psychophysics. pages 527–532. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 1994.

[18] John Watkinson. *The Art of Digital Video.* Focal Press, Oxford, England, 1995.

[19] Fung D. Y. C. and Matthews R. F. *Instrumental Methods for Quality Assurance in Foods.* Marcel Dekker Inc., New York, 1991.

[20] Fung D. Y. C. and Matthews R. F. *Instrumental Methods for Quality Assurance in Foods.* Marcel Dekker Inc., New York, 1991.

[21] Batchelor B. G., Hill D. A., and Hodgson D. C. *Automated Visual Inspection.* IFS(Publications) Ltd., UK, 1985.

[22] Drury C. G. and Fox J. G. F. *Human Reliability in Quality Control.* Halstead Press, New York, 1975.

[23] Alex Pentland and Tanzeem Choudhury. Face recognition for smart environments. *IEEE Computer*, pages 52–55, February 2000.

[24] Roderick McDonald. *Color Physics for Industry.* Society of Dyers and Colourists, West Yorkshire, England, 1987.

[25] Purves D., Augustine G. J., Fitzpatrick D., Katz L. C., LaMantia A., and McNamara J. O. *Neuroscience.* Sinauer Associates, Inc., Sunderland Massachusetts, 1997.

[26] Wandell Brian A. *Foundations of Vision.* Sinauer Associates Inc., Sunderland Massachusetts, 1995.

[27] Peter Gouras. *Vision and Visual Dysfunction: The Perception of Colour*, volume 6. CRC Press, Inc., Boston, MA, 1991.

[28] Roderick McDonald. *Color Physics for Industry.* Society of Dyers and Colourists, West Yorkshire, England, 1987.

[29] Brian A. Wandell. *Foundations of Vision.* Sinauer Associates, Inc., Sunderland, Massachusetts, 1995.

[30] Bernd Jahne, Horst Haussecker, and Peter Geissler, editors. *Handbook of Computer Vison and Applications*, volume 1. Academic Press, New York, 1999.

[31] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing.* Addison-Wesley Publishing, New York, 1992. General Image Processing Text.

[32] Author R. Weeks Jr. *Fundamentals of Electronic Image Processing.* SPIE Press, Bellinghan, Washington USA, 1996.

[33] Gregory C. DeAngelis, Izumi Ohzawa, and Ralph D. Freeman. Receptive-field dynamics in the central visual pathways. *Trends in Neuroscience*, 18(10):451–458, 1995.

[34] Audie G. Leventhal. *The Neural Basis of Visual Function*, volume 4 of *Vision and Visual Dysfunction*. CRC Press, Inc., Boston, MA, 1991.

[35] Gordon S. G. Beveridge and Robert S. Schechter. *Optimization: Theory and Practice*. McGraw-Hill, New York, 1970. Optimization text.

[36] Christina Enroth-Cugell and John Robson. Functional characteristics and diversity of cat retinal ganglion cells. *Investigative Ophthalmology and Visual Science*, 25:250–267, 1984.

[37] Theodoridis S. and Koutroumbas K. *Pattern Recognition*. Academic Press, Boston Massachusetts, 1999.

[38] Britton D. F., Daley W. D., and Galloway B. Vision-based citrus inspection and grading system. Proceedings of the Global International Signal Processing Conference, Dallas, TX, March 2003.

[39] Ricki Lewis. Probing the amazing human retina. *Biophotonics International*, pages 40–41, May/June 1999.

[40] Berthold Klaus and Paul Horn. *Robot Vision*. The MIT Press McGraw-Hill Book Company, Cambridge, Massachusetts, 1986.

Vita

Wayne Dwight Roomes Daley was born to Mr. and Mrs. Egbert Daley on October 11, 1957, in Kingston Jamaica. He received his B.S. degree in Mechanical Engineering in 1980 and an M.S. in Mechanical Engineering in 1982 from the Georgia Institute of Technology.

Wayne has worked at the Georgia Tech Research Institute as a Research Engineer since December 1982. His research interests has spanned investigations in heat transfer in pourous media to alternative energy systems. He then worked in the areas of computerized control and automation. Currenty his activities revolve around visual sensing for process monitoring, machine guidance and control and approaches towards the development of machine vision algorithms. He is also currently a Ph.D candidate in the George Woodruff School of Mechanical Engineering at Georgia Tech.