Image statistics and the perception of surface reflectance

by

Lavanya Sharan

B. Tech. Electrical Engineering
Indian Institute of Technology, Delhi, 2003

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2005

Author .....
Department of Electrical Engineering and Computer Science
August 31, 2005

Certified by. .....
Edward H. Adelson
Professor of Vision Science
Thesis Supervisor

Accepted by. .
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Image statistics and the perception of surface reflectance

by

Lavanya Sharan

Submitted to the Department of Electrical Engineering and Computer Science
on August 31, 2005, in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering and Computer Science

Abstract

Humans are surprisingly good at judging the reflectance of complex surfaces even when the surfaces are viewed in isolation, contrary to the Gelb effect. We argue that textural cues are important for this task. Traditional machine vision systems, on the other hand, are incapable of recognizing reflectance properties. Estimating the reflectance of a complex surface under unknown illumination from a single image is a hard problem. Recent work in reflectance recognition has shown that certain statistics measured on an image of a surface are diagnostic of reflectance.

We consider opaque surfaces with medium scale structure and spatially homogenous reflectance properties. For such surfaces, we find that statistics of intensity histograms and histograms of filtered outputs are indicative of the diffuse surface reflectance. We compare the performance of a learning algorithm that employs these image statistics to human performance in two psychophysical experiments. In the first experiment, observers classify images of complex surfaces according to the perceived reflectance. We find that the learning algorithm rivals human performance at the classification task. In the second experiment, we manipulate the statistics of images and ask observers to provide reflectance ratings. In this case, the learning algorithm performs similarly to human observers. These findings lead us to conclude that the image statistics capture perceptually relevant information.

Thesis Supervisor: Edward H. Adelson
Title: Professor of Vision Science

Acknowledgments

I would like to thank my advisor, Ted Adelson, for his invaluable guidance. I am grateful for his wisdom, encouragement and patience.

Yuanzhen Li, who collaborated in the investigation into diagnostic image statistics, provided many useful discussions and valuable feedback. I would also like to thank Shin'ya Nishida and Isamu Motoyoshi (NTT Communication Science Labs), Roland Fleming (Max Planck Institute for Biological Cybernetics) and Ruth Rosenholtz (MIT) for helpful discussions.

All members of the Perceptual Science group contributed to creating a very friendly work environment. Special thanks are due to John Canfield, for helping me with all my hardware and software problems in the past two years.

I would like to acknowledge our sponsors - NTT Communication Science Laboratories, Japan and NSF.

Finally, I would like to thank my parents, Basabi Bhaumik and Pankaj Sharan, for their love and support.

Contents

1 Introduction **21**

1.1 Motivation 23

1.2 Thesis Outline 25

2 Previous Work **29**

2.1 Background 29

2.2 Computer Graphics and Vision Approaches 30

2.3 Reflectance Estimation in Human Vision 34

2.4 Summary 36

3 The Role of Image Statistics **37**

3.1 Image Data Set 38

3.2 Mean luminance normalization 42

3.3 Obtaining Ground Truth 43

3.4 Statistics of Intensity Histograms 44

3.5 Statistics of Filtered Images 48

3.6 Interdependence of the statistics 58

4 Synthesizing material appearance **61**

4.1 Previous Work 62

4.2 Luminance Histogram Equalization 64

4.3 Heeger-Bergen applied to material images 67

4.4 Activity Map based Heeger-Bergen 73

5 Comparison of Image Statistics with Human Performance **87**

5.1 Observers 89

5.2 Apparatus 89

5.3 Stimuli 90

5.4 Methods 94

5.5 Experiment I 95

5.5.1 Motivation 95

5.5.2 Procedure 96

5.5.3 Results 97

5.6 Experiment IIA 102

5.6.1 Motivation 102

5.6.2 Procedure 103

5.6.3 Results 103

5.7 Experiment IIB 116

5.7.1 Motivation 116

5.7.2 Procedure 116

5.7.3 Results 117

5.8 Experiment IIC 123

5.8.1 Motivation 123

5.8.2 Procedure 123

5.8.3 Results 124

5.9 Comparison of Experiment IIc responses with Statistics 137

6 Summary **149**

6.1 Thesis Contributions 150

6.1.1 Shadows, Interreflections and Surface Mesostructure 150

6.1.2 Image Statistics and Reflectance Estimation 150

6.1.3 Psychophysics with Complex Stimuli 151

6.2 Future Work 151

6.2.1 Relaxing constraints on illumination and reflectance 151

6.2.2 Synthesizing material appearance 152

A Experiment II results **155**

B Analysis of Experiment II **189**

B.1 Experiment IIA 189

B.2 Experiment IIC 194

List of Figures

1-1 The Gelb effect - reproduced from [56] 25

1-2 The stucco samples are normalized to have the same mean luminance, yet it is easy to tell the white stucco from black. 26

3-1 Materials in Data Set 40

3-2 Lighting conditions for Image Data Set 41

3-3 Orange materials provide three images of distinct reflectance values . 41

3-4 Examples of challenging materials 41

3-5 Multiple Exposure Imaging 42

3-6 Examples of mean normalized images 43

3-7 Acquiring Ground Truth for Reflectance 44

3-8 Intensity histograms of light and dark materials exhibit systematic differences 47

3-9 Ground Truth versus Statistics of the log(Luminance) Light 1 49

3-10 Interpreting Scatter Plots 50

3-11 Reflectance vs Std. Deviation of log(Luminance) in Light 1 (top), Light 2 (middle) and Light 3(bottom) 51

3-12 Center Surround Filtering 52

3-13 The statistics of filtered outputs are diagnostic of reflectance 53

3-14 Reflectance vs Std. Deviation of filter output in Light 1 54

3-15 Reflectance vs $(90^{th} - 10^{th})$ Percentile of filter output in Light 1 . . . 55

3-16 ROC Curves 56

3-17 Examples of misclassified materials 57

3-18 Interdependence of Statistics, Standard Deviation of log(Luminance) vs ($90^{th} - 10^{th}$) Percentile of gaussian center surround 59

4-1 Pseudo Code for Heeger-Bergen reproduced from [25] 63
4-2 Pseudo Code for Match-Histogram procedure in Figure 4-1 (reproduced from [25]) 63
4-3 More successful example of histogram equalization 65
4-4 Less successful example of histogram equalization 66
4-5 An Example of Failure with Heeger Bergen 69
4-6 Direct Heeger Bergen search succeeds 70
4-7 Example of success with Heeger Bergen 71
4-8 Heeger Bergen with subsampled Laplacian pyramid 71
4-9 Another example of Heeger Bergen with subsampled Laplacian pyramid 72
4-10 Activity Map Based Heeger-Bergen 74
4-11 One Iteration of Activity Map based Heeger-Bergen, Images 75
4-12 One Iteration of Heeger-Bergen, Histograms 76
4-13 KL Distance vs Iterations 77
4-14 Comparison of methods 78
4-15 Comparison of methods (Blowup) 79
4-16 Material 2, Activity Map based Heeger-Bergen 80
4-17 Material 3, Activity Map based Heeger-Bergen 81
4-18 Material 5, Activity Map based Heeger-Bergen 82
4-19 Material 9, Activity Map based Heeger-Bergen 83
4-20 Material 14, Activity Map based Heeger-Bergen 84
4-21 Material 1 and 10, Activity map based Heeger Bergen 85

5-1 Examples of failure of image statistics 88
5-2 Setup for Experiment II 91
5-3 Munsell scale box setup 92
5-4 What the subjects viewed 93
5-5 Performance of observers and classifier in Experiment I 100

5-6 Experiment I: Errors by observers and classifier 101
5-7 Stimuli for Experiment IIA 104
5-8 Log_{10}(Perceived reflectance) vs Log_{10}(Background luminance) (Material 10, R channel) 106
5-9 Log_{10}(Perceived reflectance) vs Log_{10}(Background luminance) (Material 10, B channel) 107
5-10 Log_{10}(Perceived reflectance) vs Log_{10}(Mean Image luminance) (Material 10, R channel) 109
5-11 Log_{10}(Perceived reflectance) vs Log_{10}(Screen luminance) (Material 10, R channel) 111
5-12 Log_{10}(Perceived reflectance) vs Log_{10}(Luminance Ratio) (Material 10, R channel) 112
5-13 Material 10 results for Experiment IIA 114
5-14 Material 5 results for Experiment IIA 115
5-15 Log_{10}(Perceived reflectance) vs Log_{10}(Patch Luminance/Background Luminance), patch luminance is held fixed 120
5-16 Log_{10}(Perceived reflectance) vs Log_{10}(Patch Luminance/Background Luminance), background luminance is held fixed 121
5-17 Log_{10}(Perceived reflectance) vs Log_{10}(Patch Luminance), luminance ratio is held fixed 122
5-18 Perceived reflectance vs Ground Truth (Group 1) 126
5-19 Perceived reflectance vs Ground Truth (Group 2) 127
5-20 Perceived reflectance vs Ground Truth (Group 3) 128
5-21 Errors from Ground Truth, Group 1 Light 1 129
5-22 Errors from Ground Truth, Group 2 Light 1 130
5-23 Errors from Ground Truth, Group 3 Light 1 131
5-24 Errors from Ground Truth versus Ground Truth 132
5-25 Stimuli for Experiment IIC, Group 1 133
5-26 Perceived Reflectance vs Channel Index (Group 1, Light 1) 134
5-27 Stimuli for Experiment IIC, Group 2 135

5-28 Perceived Reflectance vs Channel Index (Group 2) 136
5-29 Training Performance ϵ-SVR, Linear Kernel, Set 1 statistics, train set Group 1 139
5-30 Comparison of ϵ-SVR with linear kernel, Set 1 statistics with averaged subject performance for training set Group 1 140
5-31 Test Performance ϵ-SVR, Linear Kernel, Set 1 statistics, train set Group 1 141
5-32 Comparison of ϵ-SVR with linear kernel, Set 1 statistics with averaged subject performance for Groups 2 and 3, train set Group 1 142
5-33 Training Performance ϵ-SVR, Linear Kernel, Set 1 statistics, train set Group 2 143
5-34 Test Performance ϵ-SVR, Linear Kernel, Set 1 statistics, train set Group 2 143
5-35 Comparison of ϵ-SVR with linear kernel, Set 1 statistics with averaged subject performance for training set Group 2 144
5-36 Comparison of ϵ-SVR with linear kernel, Set 1 statistics with averaged subject performance for Groups 2 and 3, train set Group 2 145
5-37 Training Performance ϵ-SVR, Linear Kernel, Set 1 statistics, train set Group 3 146
5-38 Test Performance ϵ-SVR, Linear Kernel, Set 1 statistics, train set Group 3 146
5-39 Comparison of ϵ-SVR with linear kernel, Set 1 statistics with averaged subject performance for training set Group 3 147
5-40 Comparison of ϵ-SVR with linear kernel, Set 1 statistics with averaged subject performance for Groups 1 and 2, train set Group 3 148

A-1 Log_{10}(Perceived reflectance) vs Log_{10}(Background luminance) (Material 10, R channel) 156
A-2 Log_{10}(Perceived reflectance) vs Log_{10}(Background luminance) (Material 10, G channel) 157

A-3 Log_{10}(Perceived reflectance) vs Log_{10}(Background luminance) (Material 10, B channel) 158
A-4 Log_{10}(Perceived reflectance) vs Log_{10}(Background luminance) (Material 5, R channel) 159
A-5 Log_{10}(Perceived reflectance) vs Log_{10}(Background luminance) (Material 5, G channel) 160
A-6 Log_{10}(Perceived reflectance) vs Log_{10}(Background luminance) (Material 5, B channel) 161
A-7 Log_{10}(Perceived reflectance) vs Log_{10}(Mean Image luminance) (Material 10, R channel) 162
A-8 Log_{10}(Perceived reflectance) vs Log_{10}(Mean Image luminance) (Material 10, G channel) 163
A-9 Log_{10}(Perceived reflectance) vs Log_{10}(Mean Image luminance) (Material 10, B channel) 164
A-10 Log_{10}(Perceived reflectance) vs Log_{10}(Mean Image luminance) (Material 5, R channel) 165
A-11 Log_{10}(Perceived reflectance) vs Log_{10}(Mean Image luminance) (Material 5, G channel) 166
A-12 Log_{10}(Perceived reflectance) vs Log_{10}(Mean Image luminance) (Material 5, B channel) 167
A-13 Log_{10}(Perceived reflectance) vs Log_{10}(Screen luminance) (Material 10, R channel) 168
A-14 Log_{10}(Perceived reflectance) vs Log_{10}(Screen luminance) (Material 10, G channel) 169
A-15 Log_{10}(Perceived reflectance) vs Log_{10}(Screen luminance) (Material 10, B channel) 170
A-16 Log_{10}(Perceived reflectance) vs Log_{10}(Screen luminance) (Material 5, R channel) 171
A-17 Log_{10}(Perceived reflectance) vs Log_{10}(Screen luminance) (Material 5, G channel) 172

A-18 Log_{10}(Perceived reflectance) vs Log_{10}(Screen luminance) (Material 5, B channel) . . . 173
A-19 Log_{10}(Perceived reflectance) vs Log_{10}(Luminance Ratio) (Material 10, R channel) . . . 174
A-20 Log_{10}(Perceived reflectance) vs Log_{10}(Luminance Ratio) (Material 10, G channel) . . . 175
A-21 Log_{10}(Perceived reflectance) vs Log_{10}(Luminance Ratio) (Material 10, B channel) . . . 176
A-22 Log_{10}(Perceived reflectance) vs Log_{10}(Luminance Ratio) (Material 5, R channel) . . . 177
A-23 Log_{10}(Perceived reflectance) vs Log_{10}(Luminance Ratio) (Material 5, G channel) . . . 178
A-24 Log_{10}(Perceived reflectance) vs Log_{10}(Luminance Ratio) (Material 5, B channel) . . . 179
A-25 Material 10 results for Experiment IIA . . . 180
A-26 Material 5 results for Experiment IIA . . . 181
A-27 Log_{10}(Perceived reflectance) vs Log_{10}(Patch Luminance/Background Luminance), patch luminance is held fixed . . . 182
A-28 Log_{10}(Perceived reflectance) vs Log_{10}(Patch Luminance/Background Luminance), background luminance is held fixed . . . 183
A-29 Log_{10}(Perceived reflectance) vs Log_{10}(Patch Luminance), luminance ratio is held fixed . . . 184
A-30 Perceived reflectance vs Ground Truth (Group 1) . . . 185
A-31 Perceived reflectance vs Ground Truth (Group 2) . . . 186
A-32 Perceived reflectance vs Ground Truth (Group 3) . . . 187

B-1 Three Factor Within Subjects ANOVA for Material 10 . . . 191
B-2 Three Factor Within Subjects ANOVA for Material 5 . . . 192
B-3 Four Factor Within Subjects ANOVA for Materials 5 and 10 . . . 193

B-4 Two Factor Within Subjects ANOVA, Keeping Background and Image Luminance Fixed 196
B-5 Errors from Ground Truth, Group 1 Light 1 197
B-6 Errors from Ground Truth, Group 1 Light 2 198
B-7 Errors from Ground Truth, Group 1 Light 3 199
B-8 Errors from Ground Truth, Group 2 Light 1 200
B-9 Errors from Ground Truth, Group 2 Light 2 201
B-10 Errors from Ground Truth, Group 2 Light 3 202
B-11 Errors from Ground Truth, Group 3 Light 1 203
B-12 Errors from Ground Truth, Group 3 Light 2 204
B-13 Errors from Ground Truth, Group 3 Light 3 205
B-14 Errors from Ground Truth, Group 1 all lights 206
B-15 Errors from Ground Truth, Group 2 all lights 207
B-16 Errors from Ground Truth, Group 3 all lights 208
B-17 Perceived Reflectance vs Channel Index (Group 1, Light 1) 209
B-18 Perceived Reflectance vs Channel Index (Group 1, Light 2) 210
B-19 Perceived Reflectance vs Channel Index (Group 1, Light 3) 211
B-20 Perceived Reflectance vs Channel Index (Group 2) 212
B-21 Perceived Reflectance vs Channel Index (Group 3, Light 1) 213
B-22 Perceived Reflectance vs Channel Index (Group 3, Light 2) 214
B-23 Perceived Reflectance vs Channel Index (Group 3, Light 3) 215

List of Tables

3.1 Reflectance Estimate for R channel of orange materials 45

3.2 Reflectance Estimate for G channel of orange materials 45

3.3 Reflectance Estimate for B channel of orange materials 46

3.4 Reflectance Estimate for white and black materials 46

3.5 Interdependence of Statistics 60

5.1 Experiment I results 99

Chapter 1

Introduction

Humans are remarkably adept at recognizing the reflective properties of surfaces. We find it easy to distinguish a shiny plastic spoon from a matte wooden spoon or a lustrous stainless steel spoon. This ability, known as material perception, is striking because the appearance of a surface varies greatly as a function of its environment. For example a stainless steel spoon reflects its surroundings, therefore images of the spoon when placed indoors on a kitchen table or outdoors on a patio table, will differ widely in a pixelwise sense. Moreover, images of a shiny plastic spoon and a steel spoon in the same environment may be more similar pixelwise than the images of each spoon in different environments.

Material recognition is different from object recognition. Object recognition involves distinguishing between different classes of objects e.g. spoons from forks or knives. Object recognition is a well studied problem while material recognition has received less attention. Template matching, commonly used to solve object recognition tasks, is not directly applicable to material recognition, as different objects may have identical material properties. Current machine vision systems lack the ability to discriminate material properties like reflectance, translucency, glossiness, wetness etc. For many vision applications such material recognition capabilities are desirable. For example a domestic robot may need to distinguish between woollen and cotton clothes while sorting soiled laundry.

Dror et al [14, 15, 16] made an important contribution to the problem of material recognition. They built a machine vision system that can identify the reflective properties of a sphere under unknown complex illumination from a single monochrome image. Their system learns the relationship between the statistics of an image and the reflective properties of a surface. By measuring relevant image statistics, the system classifies surfaces as shiny, matte, white, gray, chrome etc. For surfaces of spatially homogenous reflective properties and known geometry, their system rivals human performance.

In this thesis, like Dror et al we identify image statistics that are diagnostic of the reflective properties of a surface. We consider opaque materials of homogenous reflective properties under simple artificial illumination. We allow our materials to possess three dimensional medium scale structure or *surface mesostructure* [29]. We assume that a single monochrome image of a flat, planar sample of the material is available. Image statistics measured on this image allow us to estimate reflectance attributes of the material. We compare the performance of our statistics with that of human observers in two experiments. In the first experiment, we asked observers to classify images of materials into two categories - light(white) or dark(black). In the second experiment, observers compare images of materials to a set of standard surfaces. The observers then indicate the standard surface whose reflectance properties are closest to those of the material. In both experiments, we find that a learning algorithm trained on our selection of image statistics performs as well as a human observer.

The work in this thesis differs from Dror et al's in that we make a different set of assumptions about the same problem. Dror et al assume that material samples are smooth spheres with no surface mesostructure. They allow complex unknown illumination as long as it is representative of the real world illumination. We assume that material samples are flat, planar patches. We allow surface mesostructure, however

we assume simple artificial lighting. We only consider opaque materials, while Dror et al permit translucency. Therefore the class of images that can be handled by Dror et al's algorithm is distinct from the class we consider in this work.

The reflective properties of a real world surface can be specified by the bidirectional reflectance distribution function or BRDF. Knowledge of the BRDF allows us to predict the appearance of the surface under any arbitrary lighting condition. The BRDF of an ideal diffuse or Lambertian surface is a constant. This constant known as the albedo or diffuse surface reflectance, is the fraction of incident light that is reflected by the surface. Real world surfaces like matte paint can be approximated as Lambertian surfaces. In this thesis we will use the term reflectance in the sense of albedo, even though most of our materials are not Lambertian. Note that we are not assuming Lambertian surfaces; we choose to quantify only the Lambertian component of the reflectance phenomena.

1.1 Motivation

Estimating the reflective properties of an isolated surface is an underconstrained problem. The luminance at a point on the surface is a function of the surface reflectance and the incident illumination. The luminance can be measured by our eyes (or by a camera) and assumed a given, however if the illumination is unknown it is not possible to solve for the reflectance. A compelling demonstration of this fact is the classic Gelb effect (Figure 1-1). A smooth white paper disc and a smooth black paper disc are suspended in a dark room and illuminated by hidden projectors. The illumination on the two discs is adjusted independently so that the luminance of the two discs is the same. A human observer who is unaware of the difference in illumination, views the discs. The observer perceives the discs to have the same reflectance, even though one of the discs is white and the other is black.

The Gelb effect is an example of the failure of *lightness constancy* - the human ability to estimate the reflectance of surfaces across a vast range of illumination and surround conditions. Perceived reflectance is known as lightness. The setup in Figure 1-1 is special and uncommon in our daily visual experience. By enforcing smooth surfaces, absence of background and equal luminance, a minimal stimulus is created which can fool the human visual system. However, we typically encounter richer visual stimuli and we do reasonably well at judging lightness in such cases. Consider the following modification to the Gelb experiment - replace the smooth discs with textured surfaces like stucco (Figure 1-2). The illumination is adjusted so that the stucco surfaces have the same *mean* luminance. A human observer will now perceive that the surfaces differ distinctly in their lightness. Thus the Gelb effect fails, indicating that textural cues are important for lightness judgements. Though the illumination is not known explicitly, textural cues provide sufficient constraints to make the reflectance estimation problem feasible. In this work, we want to quantify such textural cues in the form of image based measurements.

The perception of lightness involves low, middle and high level visual processing [1, 23]. There is evidence that image features like contours, junctions and brightness distributions are used by mid level vision mechanisms to deduce lightness. Adelson [1] has proposed a statistical estimation framework for lightness perception. The visual system knows (or learns) the mapping from the reflectance of a surface to the luminance distributions produced when it is placed under various lighting conditions. Given the mapping, the visual system can optimally estimate the reflectance of a new, previously unseen surface. Thus examining the role of image based statistics in reflectance estimation contributes to our understanding of the perceptual mechanisms of lightness perception.

The main application of our work is building machine vision systems that can recognize materials. Domestic or industrial robots require material recognition capabilities for many tasks. For example a domestic robot may need to tell a metallic

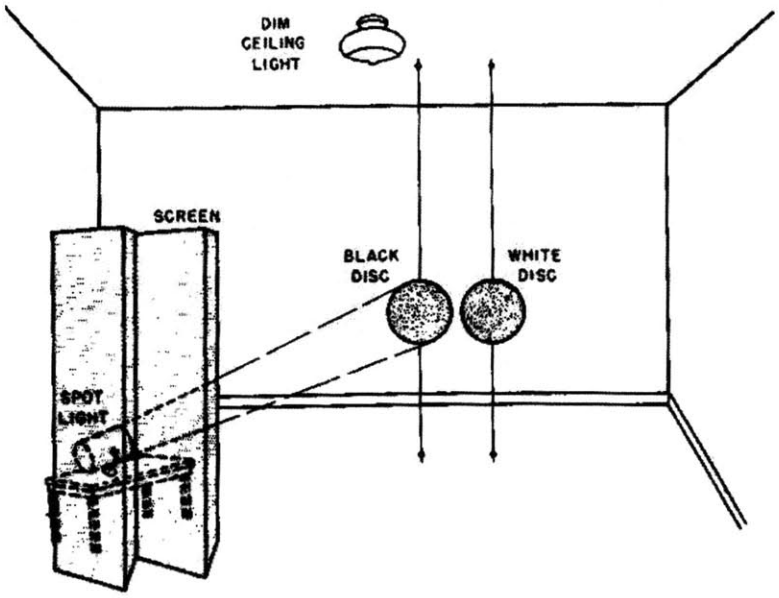


Fig. 12. Concealed illumination experiment. The person who does not know about the spotlight will judge the black and the white discs to have equal brightness. A small piece of paper thrust into the light beam, however, makes it obvious that the left disc is black.

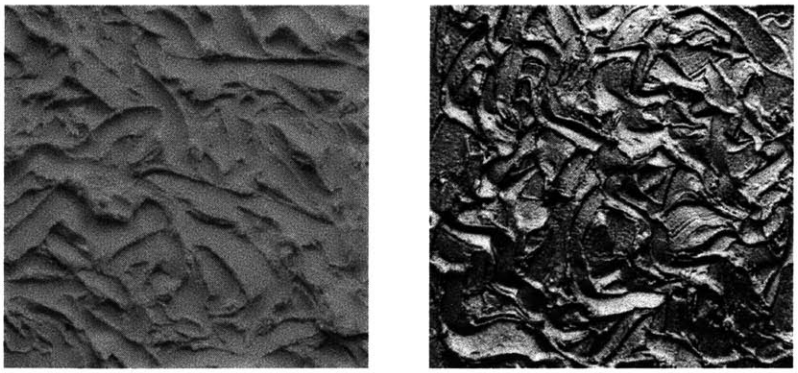
Figure 1-1: The Gelb effect - reproduced from [56]

plate from a non-metallic plate in order to heat food in the microwave. An industrial robot that performs product inspection may have to recognize the correct material in order to judge a finished product.

Understanding the interaction of surface reflectance, geometry, illumination and the resulting image, can enhance machine vision algorithms like motion estimation, or shape estimation. Such an understanding would also aid computer graphics applications such as recovering the geometry of a real world scenes from their photographs.

1.2 Thesis Outline

In next chapter, we discuss background material and previous work in the computer graphics, vision and human vision communities. We define BRDFs and describe im-



(a) White Stucco (b) Black Stucco

Figure 1-2: The stucco samples are normalized to have the same mean luminance, yet it is easy to tell the white stucco from black.

age based techniques for measuring them. We relate our work to existing research in image based reflectance estimation and lightness perception.

In Chapter 3, we begin by describing how we obtain our data set of images of materials and the ground truth for the diffuse reflectance. Next, we discuss image based measures like moment and percentile statistics of luminance histograms and histograms of filtered outputs. We observe that these statistics are diagnostic of the diffuse reflectance of our materials.

In Chapter 4, we manipulate such diagnostic image statistics to produce changes in perceived reflectance of materials. We discuss prior work in texture analysis and synthesis. We introduce a modification to the Heeger-Bergen texture synthesis algorithm to synthesize material appearance.

In Chapter 5, we compare the performance of our image statistics at estimating reflectance with human observers in two experiments. In the first experiment, a two alternative forced choice (2AFC) design, observers are asked to classify images of materials as light or dark. In the second experiment, a rating task, observers compare images of materials to a set of standard Munsell patches. The observers indicate the patch whose reflectance properties are closest to those of the material. In both exper-

iments, we find that a learning algorithm trained on our selection of image statistics performs comparably to a human observer.

In Chapter 6, we summarize the contributions of our work and outline directions for further research.

Chapter 2

Previous Work

2.1 Background

The reflectance properties of a surface are defined by its bidirectional reflectance distribution function or BRDF. The BRDF specifies what proportion of the light incident on an infinitesimal surface patch in any direction is reflected in any other direction. It is a function of four variables - (θ_i, ϕ_i) and (θ_r, ϕ_r) - the spherical coordinates of the direction of incident and reflected light respectively. The BRDF is defined for an infinitesimal patch so it can vary from point to point on a surface. Materials with homogenous reflectance properties have spatially uniform BRDFs. The reflectance properties of a surface also depend on the wavelength of light. The wavelengths of the incident and reflected light can be incorporated into the BRDF by two additional variables - λ_i and λ_r For most materials λ_i is the same as λ_r so we only need one additional variable λ. However, some materials exhibit fluorescence, a phenomena where λ_i is distinct from λ_r. In this work, we will ignore fluorescence effects.

The BRDF formulation assumes an opaque surface. It is possible to extend this formulation to include the effect of translucency. Translucent materials like skin, wax, soap etc. scatter light in addition to reflecting it. The light incident at a point of the surface undergoes sub-surface scattering and emerges at another point close-by. The bidirectional scattering-surface reflectance distribution function or the BSSRDF

captures this effect. An extension of the BRDF, it depends on eight variables - the incident as well as exitant locations and directions of light.

At this point, it is useful to clarify the relationship between surface reflectance and surface texture. Texture may result from variations in reflectance properties (wallpaper type) or from variations in fine-scale surface geometry (3D type) [40]. A surface may be described as a 3D texture at a scale where the surface roughness can resolved visually or by a camera. The appearance of a texture especially the 3D type, alters dramatically with change of lighting and viewing position. These variations are captured by the bidirectional texture function or BTF [10]. For all combinations of lighting and viewing direction, the BTF specifies the two-dimensional image (photograph) of the visible texture. The materials we use in this work may be classified as 3D textures with spatially uniform reflectance properties.

2.2 Computer Graphics and Vision Approaches

The BRDF and its variants are of tremendous importance in the field of computer graphics. Knowledge of the full BRDF permits realistic renderings of materials and objects in synthetic scenes. As the BRDF is a function of four or more variables, the space of all physically realizable BRDFs is vast. BRDFs have been approximated by parametric models to allow efficient rendering algorithms [39, 24, 54]. Such models are derived from the optics of surface reflection or are fitted to observed BRDF data. Most models distinguish between two aspects of the reflectance phenomenon - the diffuse and specular reflectance.

An ideal diffuse or a Lambertian surface reflects light uniformly in all directions regardless of the direction of incident light. Therefore the appearance of a diffuse surface is independent of the viewing direction. Diffuse reflection is caused when light undergoes multiple scattering within a surface and emerges in a random direction.

The BRDF of a Lambertian surface equals a constant known as the albedo. The albedo is the fraction of incident light energy that is reflected by the diffuse surface.

Real world diffuse surfaces such as matte paint, paper, plaster etc. depart from ideal diffuse behavior. The net direction of reflected light is sensitive to the surface geometry and number of inter-reflections therefore exhibits a directional dependence. Several authors [36, 28] have analyzed such non-Lambertian diffuse surfaces and developed BRDF formulations that lead to visually pleasing renderings of many real world diffuse surfaces.

Specular reflection is observed in smooth surfaces such as mirrors, polished metal etc. A collimated beam of light incident at point of an ideal specular surface is reflected in a direction determined by the laws of reflection, producing a sharply defined reflected beam. Therefore the appearance a specular surface depends greatly on the viewing direction. Real world specular surfaces however are never perfectly smooth, thus the reflected beam is not collimated but has an angular spread about the ideal direction of reflection. Both ideal specular and diffuse behaviors are extremes of a continuum of reflection modes. Most real world materials display both specular and diffuse reflection properties.

While parameterized models represent the reflectance properties of several common materials effectively, they fail to capture a range of real world reflectance phenomena. For such cases, empirically measured BRDFs can be used instead of parameterized models. Traditionally BRDFs are measured by a gonioreflectometer, a device where a surface sample is illuminated by a movable point light source and the reflected light is measured at all viewing angles. BRDF acquisition by a gonioreflectometer is usually a time consuming process. As an alternative a number of image based BRDF estimation techniques have been developed [44, 58, 30, 57, 52, 11, 5, 35, 42, 12]. These techniques estimate the BRDF from photographs of a surface.

To understand the image based BRDF estimation problem better, consider the following example by Marschner et al [30]. Say we want to recover the BRDF of an isotropic material i.e. a material where the reflectance depends only on the relative direction of the incident and exitant light rays. In that case the BRDF is function of $\Delta\phi = \phi_r - \phi_i$ rather than of both ϕ_i and ϕ_r. Let us ignore the dependence of the BRDF on the wavelength λ for now. The BRDF we wish to estimate is a function of three variables $(\theta_i, \theta_r, \Delta\phi)$. To recover the BRDF completely we must satisfy three degrees of freedom. A gonioreflectometer setup comprises a flat surface sample, a movable light source and a detector. By allowing all relative positions of the source and the detector, the three degrees of freedom are fulfilled and the BRDF can be measured entirely.

An equivalent solution is obtained by fixing the light source, rotating the surface sample along two orthogonal axes and moving the detector along a curve. Thus the total degrees of freedom is still three and all configurations of the light source, detector and sample can be achieved. One image based solution to the BRDF estimation problem proposed by Marschner et al [30] is as follows - the flat surface sample is replaced by a curved surface and the detector is replaced by a camera. The light source is fixed and the camera moves along a curve (one dof) taking images of the surface sample. Each two dimensional image of the curved space contributes two degrees of freedom and one acquires the same measurements as the previous solutions. By replacing a flat surface by a curved surface one avoids the physical rotation of the sample. Moreover, by using the camera, two degrees of freedom are sampled in parallel reducing the time complexity of the BRDF measurement.

To summarize, arbitrary BRDFs are functions of four or more variables. In order to measure a BRDF completely, the appropriate degrees of freedom must be satisfied by any measurement technique. Image based techniques accelerate BRDF measurement but since an image only contributes two degrees of freedom, additional information or equipment is required to achieve the total degrees of freedom. For

the Marschner et al technique described above, this implies sufficient curvature of the sample and multiple images. Image based techniques differ in how they resolve the extra degrees of freedom. Some require additional equipment such as a laser range scanner or a light stage, or human interaction or prior knowledge of the geometry or illumination.

Our formulation of the reflectance estimation problem - using a single image under unknown but simple illumination and arbitrary surface mesostructure is severely underconstrained for full BRDF estimation. Several combinations of geometry, illumination and reflectance can explain the given image. However, one interpretation of the image is more *likely* to occur in the real world than others. Humans when placed in similar conditions (isolated viewing of a surface under unknown illumination) extract such an interpretation unconsciously. This suggests that our problem is not as underconstrained as it appears at first glance.

The reflectance estimation problem can be viewed in a probabilistic framework. The most likely interpretation of an image may be obtained by integrating the posterior probability of each reflectance function over all possible illuminations [21]. Such a Bayesian formulation requires the specification of the prior probability distribution. While real world illuminations exhibit statistical regularities [15] an explicit probability distribution over all illuminations is hard to specify. Nevertheless, it is possible to decompose an image into intrinsic "illumination" and "reflectance" images [55, 49, 50]. Weiss [55] decomposed a sequences of images into intrinsic images by assuming a prior that illumination images result in sparse filter outputs. Tappen et al [49, 50] use local information from the color and intensity patterns in an image to separate a single color image into intrinsic "shading" and "reflectance" images.

To sum up, estimating the BRDF of a surface from a single image under unknown illumination seems impossible. However since humans can estimate the reflective properties of a surface under similar conditions, we believe that there is sufficient

information in the image to allow accurate reflectance estimation. Such image information may be quantified as statistics of intensity distributions or distributions of filter outputs. In this work, we do not attempt to estimate an arbitrary BRDF. We simplify our problem by considering opaque materials with spatially uniform BRDFs. Furthermore we allow our materials to be non-Lambertian but we focus mainly on the diffuse reflectance.

Like Dror et al [16, 14] we do not estimate the illumination or the surface structure explicitly. Instead we search for image statistics that are diagnostic of diffuse reflectance across variations in illumination and surface mesostructure. We can build a reflectance estimator using machine learning techniques that can learn the relationship between diagnostic statistics and diffuse reflectance. As we noted earlier in Chapter 1, our work differs from Dror et al's in that the class of images handled by their algorithm is distinct from the class of images we consider.

2.3 Reflectance Estimation in Human Vision

As mentioned in Section 1.1, lightness constancy forms the main motivation for this research. Humans can estimate some of the reflectance properties of complex surfaces, even in a single isolated viewing. This ability is impressive given the underconstrained nature of the reflectance estimation problem. Lightness constancy is a well-studied problem in the field of lightness perception.

Historically there have been two approaches to lightness constancy - the low level and high level approaches. Hering [27] proposed that low level physiological mechanisms like adaptation and local interactions are critical for lightness constancy. Helmholtz [26] on the other hand, proposed a high level vision approach whereby our visual system, based on past experience, *infers* the most probable estimate of lightness. Recent psychophysical studies have found evidence for mid level vision

mechanisms based on image features like contours, junctions and local brightness distributions [1, 48, 51, 59, 23].

Most work on lightness perception has focused on diffuse reflection from flat Lambertian surface patches under artificial illumination. Such conditions are uncommon in our daily visual experience; we normally encounter non-Lambertian surfaces under complex real world illumination. Recently a number of studies have focused on stimuli that incorporate some of the complexity of real world conditions.

Nishida and Shinya [34] conducted psychophysical experiments to measure the accuracy of human surface reflectance estimation. They found that observers fail to estimate the reflectance of surfaces of arbitrary shape under point source illumination. They showed that the observers' matches correlate strongly with the luminance histograms of the images. Fleming et al [19] showed that observers can estimate the reflectance of a surface accurately when the illumination is representative of that found in the natural world scenes. This suggests that humans implicitly use statistics of real world illumination to estimate reflectance.

There have been a number of studies on the non-Lambertian aspects of reflectance perception such as gloss. Beck and Prazdny [4] demonstrated that the perception of surface gloss may involve low and mid level visual cues. Pellacini et al [37] reparametrized the space spanned by the Ward reflection model to create a perceptually uniform gloss space. Robilloto and Zaidi [43] measured the limits of lightness identification for real objects in a 3D setup under natural viewing conditions. They found that observers seem to use brightness dissimilarity to judge lightness.

Recent work in our group [2, 3, 45] and by Motoyoshi et al [32, 31, 33] demonstrates that simple image based statistics are indicative of surface reflectance. Motoyoshi et al have shown that moment statistics of the luminance histogram and subband histograms of images of real world textured surfaces are correlated with the perceived

reflectance and gloss of a surface.

In this work, we conduct psychophysical studies to explore the accuracy of lightness identification for images of real world surfaces. We display the images on an LCD display and normalize them to have the same mean luminance. We find that the Gelb effect fails and that observers can, to some extent, estimate the reflectance in the absence of mean luminance information. We find that the performance of a learning algorithm that uses informative image statistics is comparable to an average human observer.

2.4 Summary

Real world surfaces display a range of reflectance properties that can be specified by the bidirectional reflectance distribution function and its variants. Recently a number of image based BRDF estimation techniques have been developed in the computer graphics community. These techniques assume known illumination or geometry or multiple photographs. Estimating the reflectance properties of real world surface from a single image under unknown illumination is underconstrained problem. However, humans seem to solve this problem under similar conditions in the real world. Therefore, we believe there is sufficient information in a single image to make the reflectance estimation problem feasible. In the next chapter, we quantify this image information in the form of statistics of image intensity distributions and distributions of filtered outputs.

Chapter 3

The Role of Image Statistics

We are influenced by the work of Fleming, Dror and Adelson [16, 19] on real world illumination statistics and lightness perception. The authors argue against the high level or the *inverse optics* approach to lightness constancy. According to this approach, the visual system can recover the reflectance of a surface by forming a precise estimate of the illumination and thereby *discount the illuminant.* In order to achieve such precise discounting, the visual system must know something about real world optics. Thus when confronted with incomplete information, such as a single image, the system can still make reasonable inferences.

Fleming et al [19, 20] advocate an alternate *measurement* approach. They propose that the visual system makes image measurements that are diagnostic of material properties but are invariant to changes in illumination. Therefore the visual system need only measure such invariant properties of an image to estimate say the reflectance. It does not have to explicitly estimate the illumination or perform inverse optics. Such image measurements, argue the authors, may be accomplished by mid level vision mechanisms.

Dror et al [15, 16] offer a quantitative form for these image measurements. They observe that real world illuminations exhibit a statistical structure similar to that of natural images. This structure can be specified in terms of statistical measures like

pixel histograms, power spectra or wavelet coefficients of illumination maps. Dror et al demonstrate the robustness of such measures to changes in illumination. They then consider images of smooth shiny spheres rendered under various real world illumination maps. The authors observe that the same statistical measures computed on the images of the sphere are diagnostic of reflectance in addition to being invariant to illumination changes. Thus certain image features or statistics may be correlated with illumination invariant material properties. It is conceivable that our visual system uses similar image based measures to deduce material properties like reflectance, gloss, translucency etc.

We want to find statistics that are diagnostic of reflectance of real world materials. We restrict ourselves to materials that are opaque and have spatially homogenous reflectance properties. We assume our materials samples are flat, planar patches however we allow surface mesostructure.

3.1 Image Data Set

We built an image data set of real world materials. We photographed several commonly encountered materials such as paper, candies, cloth and several hand made surfaces such as stucco (see Figure 3-1). Our data set contains 30 materials in three different lighting conditions. The images were acquired in a RAW 12-bit format by a Canon EOS 10D camera. The RAW images were linearized using Dave Coffin's dcraw [9] software. In this work we are primarily concerned with lightness perception and not color therefore we convert our color image data to gray scale.

The three lighting conditions will be referred to as Light 1, 2 and 3. Light 1 was an overhead fluorescent light (Kino Flo Diva Lite 200). Light 2 was a halogen spotlight (LTM Pepper 300 W Quartz-Fresnel Light). Light 3 was a diffuse soft light source (Lowel Rifa Lite 66, 750 W Tungsten Halogen lamp). Figure 3-2 shows photographs

of materials under the different lights.

We photographed several black, white, orange, red and yellow materials. We chose orange (and red and yellow) materials because the red channel of an orange object looks like a white or light gray material and the blue channel like that of a dark gray or black material (see Figure 3-3). Thus for each material we can obtain a black-white or light-dark pair of images, allowing us to study the effect of reflectance independently of the geometry or illumination. It is easy to achieve such a separation with synthesized images using computer graphics packages, however for real world photography it is hard to do so even in a controlled laboratory setting.

Many of our materials such as the shiny TicTacs (refer Figure 3-1) have strong specular highlights. In order to capture such materials with a limited dynamic range camera, we used the technique of multiple exposure imaging. The multiple exposures are combined using HDRShop software (Version 2) [13] to produce a single high dynamic range image.

The range of reflectance phenomena observed in real world materials is fairly daunting (Figure 3-4). We chose to limit the samples in our data set to opaque materials with spatially uniform reflectance. We do so for two reasons. One, our formulation of the reflectance problem is acutely underconstrained, hence we would like to restrict the solution space of reflectance functions. Two, we intend to conduct psychophysical experiments with our image data. It is hard for subjects to make judgements about the reflectance of materials like feathers or translucent materials like jelly beans in Figure 3-4.

Figure 3-1: Materials in Data Set

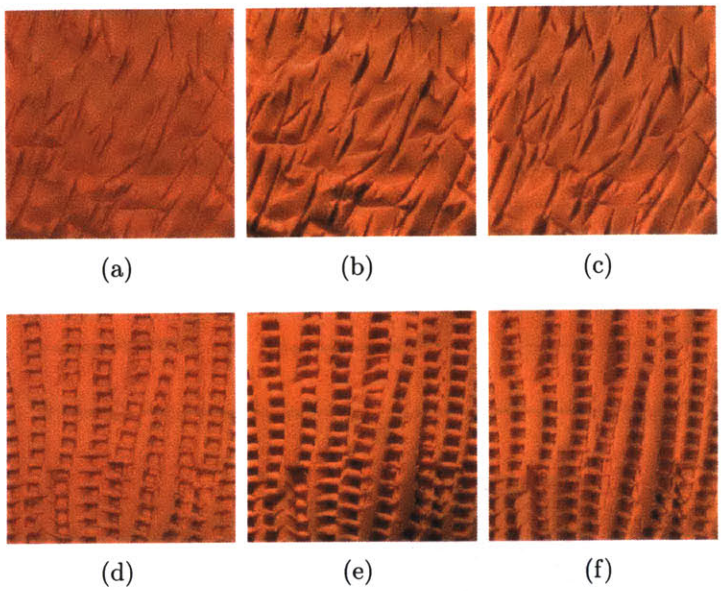


Figure 3-2: Three different lighting conditions were used. (a),(d) Overhead fluorescent light (b), (e) Focused Halogen Spotlight (c),(f) Diffuse Halogen Light

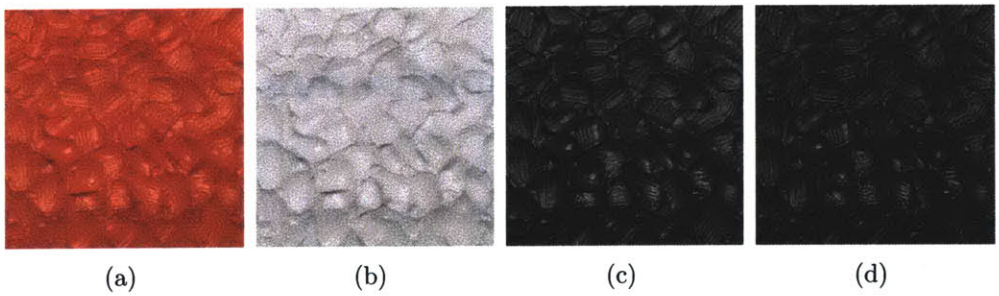


Figure 3-3: Each color channel of an orange material has a distinct reflectance (a) Orange image (b) R channel (c) G channel (d) B channel

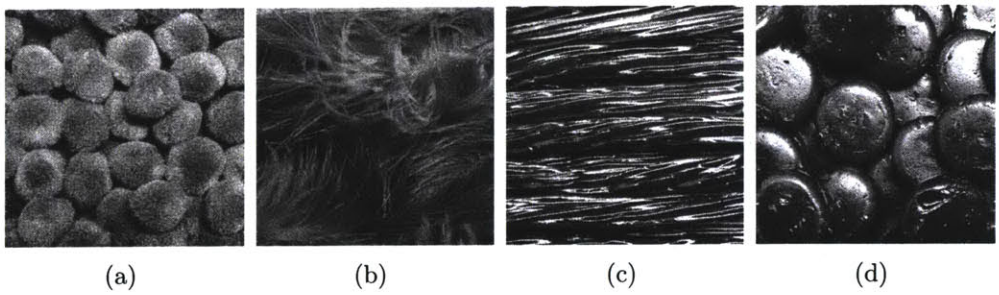


Figure 3-4: Examples of challenging materials

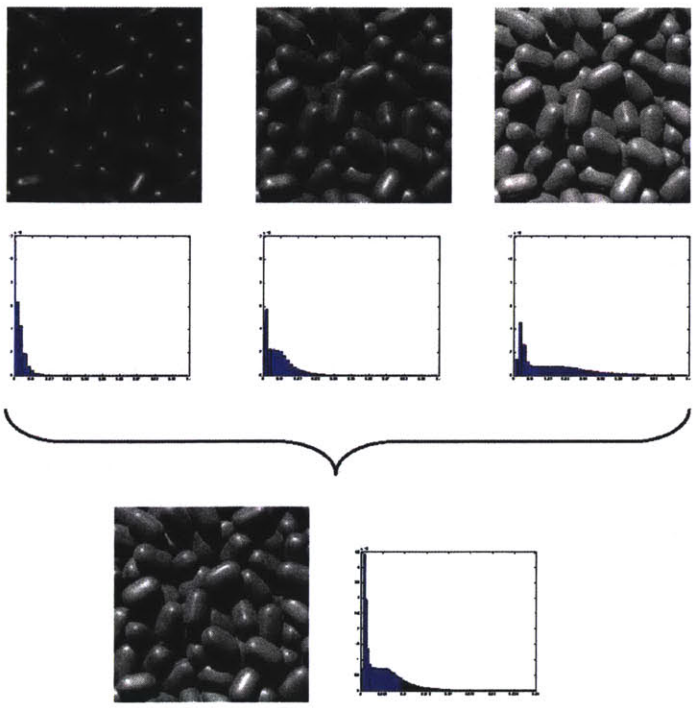
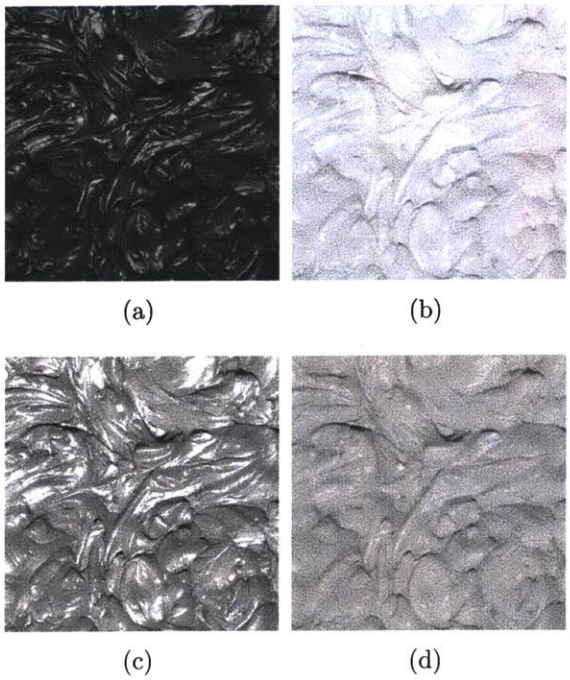


Figure 3-5: Multiple Exposure Imaging

3.2 Mean luminance normalization

In this work, we focus on recovering the reflectance from single images of materials. As the context of these images is not known, it is impossible to separate the overall level of illumination from the reflectance. The Gelb effect, described in Chapter 1, is an example of such an ambiguity. We compensate for this ambiguity by normalizing the mean luminance of all our images.

Mean luminance is an important cue for lightness perception, especially when there is contextual information. However in the absence of context, as in the anti-Gelb experiment (Figure 1-2), humans might use other cues such as texture to estimate the lightness. Therefore, by analyzing the statistics of mean normalized images we hope to quantify such textural cues.



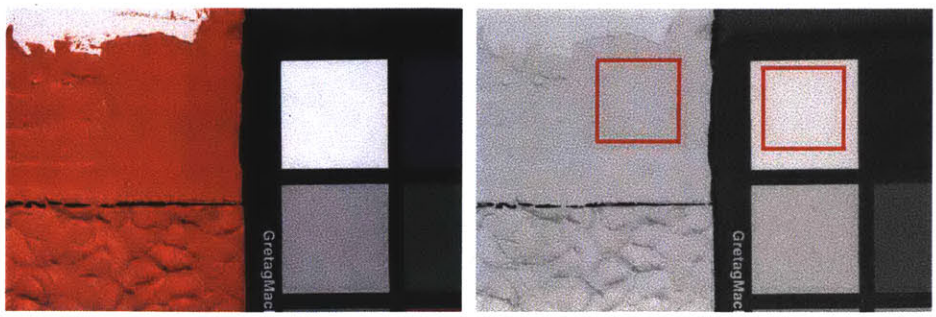
(a) (b)

(c) (d)

Figure 3-6: Examples of mean normalized images

3.3 Obtaining Ground Truth

In order to conduct meaningful psychophysical tasks with this data set it is essential to acquire the ground truth for reflectance. We obtained the ground truth for our materials by the following procedure - a smooth flat sample (i.e. without surface mesostructure) of the material was positioned next to a standard white surface (Gretag-MacBeth Color Checker Chart). The material and the standard surface were illuminated by two lamps (SunWave Full Spectrum 5500 K fluorescent bulbs). The luminance at all points on the sample and the surface was measured with lightmeters (Sekonic L-608 and Minolta CS-100 Chromameter) to ensure uniform illumination. The material sample was then photographed at multiple exposures. The image data was linearized using the workflow described in Section 3.1. Next, for each exposure the ratio between the mean luminance of a region containing the material sample and a region containing the standard surface was calculated (Figure 3-7). As the reflectance of the standard surface is known, the ratio of mean luminances allowed us to estimate the reflectance of the material sample by the following formula.



(a) (b)

Figure 3-7: **Acquiring Ground Truth** (a) Uniformly illuminated material sample and standard white surface (b) R channel of image in (a). User clicks two regions, one on the material sample and one on the standard surface. Ratio of means of the regions is used to calculate the reflectance of the material in the R channel.

$$Reflectance_{Material} = \frac{MeanLuminance_{Material}}{MeanLuminance_{Standard}} * Reflectance_{Standard} \quad (3.1)$$

For this formula to be applicable it is essential that the images be linear and that the material and standard surface receive the same illumination. It is important for the material sample and the standard surface to have the same geometry. In our case, both the sample and the standard surfaces were completely flat and planar. The reflectance estimate thus obtained was averaged across multiple exposures. We found that reflectance estimates do not vary too much across exposures (see Tables 3.1,3.2,3.3 and 3.4). For orange materials, the calculation described above, was repeated for each of the channels - R,G and B. In the final averaging step a different set of exposures is chosen for each channel as it is hard to acquire a single image where all three color channels are well exposed.

3.4 Statistics of Intensity Histograms

Motivated by Nishida and Shinya's [34] and Dror et al's [14] findings, we examined the pixel intensity histograms of our materials. We find that intensity histograms

Material	Mean Estimate	Std. Deviation	95% CI	
			Upper Limit	Lower Limit
1	0.6786	0.0022	0.6768	0.6804
2	0.7479	0.0016	0.7466	0.7492
3	0.7690	0.0031	0.7665	0.7714
4	0.4684	0.0041	0.4652	0.4717
5	0.6094	0.0013	0.6084	0.6105
6	0.6532	0.0043	0.6498	0.6566
7	0.7054	0.0038	0.7024	0.7084
8	0.6236	0.0024	0.6217	0.6255
9	0.6652	0.0008	0.6645	0.6658
10	0.6729	0.0012	0.6719	0.6739
11	0.4186	0.0014	0.4174	0.4197
12	0.6972	0.0008	0.6965	0.6978
15	0.5061	0.0015	0.5049	0.5073
17	0.6094	0.0013	0.6084	0.6105
37	0.6844	0.0013	0.6833	0.6854
38	0.4801	0.0008	0.4795	0.4808

Table 3.1: Reflectance Estimate for R channel of orange materials

Material	Mean Estimate	Std. Deviation	95% CI	
			Upper Limit	Lower Limit
1	0.2018	0.0019	0.2003	0.2033
2	0.3180	0.0015	0.3168	0.3192
3	0.4152	0.0016	0.4139	0.4165
4	0.1144	0.0009	0.1136	0.1151
5	0.2549	0.0012	0.2539	0.2559
6	0.1816	0.0041	0.1783	0.1848
7	0.2939	0.0008	0.2932	0.2945
8	0.2076	0.0016	0.2063	0.2089
9	0.3599	0.0018	0.3584	0.3613
10	0.1711	0.0037	0.1681	0.1740
11	0.1204	0.0010	0.1197	0.1212
12	0.2542	0.0040	0.2510	0.2574
15	0.1867	0.0030	0.1843	0.1891
17	0.2549	0.0012	0.2539	0.2559
37	0.2292	0.0011	0.2284	0.2301
38	0.1698	0.0007	0.1692	0.1704

Table 3.2: Reflectance Estimate for G channel of orange materials

Material	Mean Estimate	Std. Deviation	95% CI	
			Upper Limit	Lower Limit
1	0.0545	0.0018	0.0525	0.0566
2	0.1666	0.0009	0.1657	0.1676
3	0.1558	0.0010	0.1546	0.1569
4	0.0822	0.0018	0.0802	0.0842
5	0.0962	0.0009	0.0952	0.0972
6	0.0788	0.0007	0.0780	0.0795
7	0.0994	0.0011	0.0981	0.1007
8	0.0870	0.0027	0.0840	0.0900
9	0.2318	0.0020	0.2295	0.2341
10	0.0574	0.0003	0.0571	0.0577
11	0.0854	0.0029	0.0821	0.0886
12	0.0921	0.0005	0.0914	0.0927
15	0.1123	0.0016	0.1105	0.1141
17	0.0962	0.0009	0.0952	0.0972
37	0.0889	0.0024	0.0862	0.0917
38	0.0843	0.0003	0.0841	0.0846

Table 3.3: Reflectance Estimate for B channel of orange materials

Material	Mean Estimate	Std. Deviation	95% CI	
			Upper Limit	Lower Limit
2	0.0514	0.0004	0.0510	0.0518
3	0.8266	0.0001	0.8264	0.8267
4	0.0494	0.0014	0.0479	0.0510
5	0.8602	0.0012	0.8588	0.8615
6	0.8813	0.0011	0.8800	0.8825
7	0.0408	0.0005	0.0402	0.0414
31	0.0517	0.0010	0.0506	0.0528
32	0.8434	0.0011	0.8422	0.8447
33	0.0517	0.0010	0.0506	0.0528
34	0.7570	0.0020	0.7547	0.7592

Table 3.4: Reflectance Estimate for white and black materials

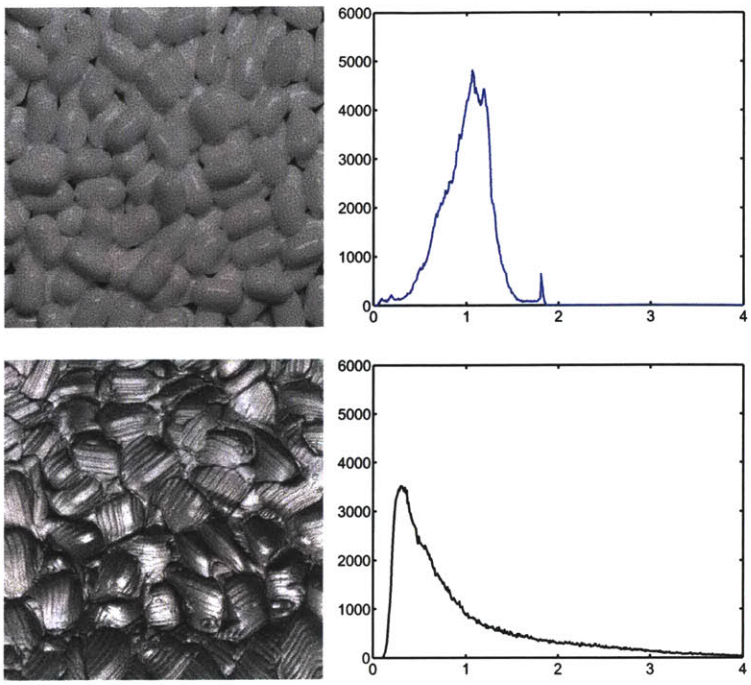


Figure 3-8: Intensity histograms of light and dark materials exhibit systematic differences

of light and dark materials show systematic differences. Consider the materials and their histograms in Figure 3-8. The histograms of the dark materials tend to have higher standard deviations and are usually positively skewed. This observation can be explained thus - materials of higher reflectance (light) have more inter-reflections hence light bounces around filling up the shadows, leading to lower local contrast as opposed to materials with lower reflectance (dark). A lower contrast translates to a lower standard deviation of the intensity histogram since the images have been normalized to equate the means.

Both light and dark materials have the same amount of specular reflection, however the contrast is higher in darker materials hence specularities are more visible, leading to longer tails and positive skew in the intensity histograms. In addition to moment statistics like standard deviation and skew, we observe that percentile statistics like 10^{th} or 90^{th} percentile or the median are useful for discriminating light materials from dark ones.

Figure 3-9 plots the reflectance estimate of our materials versus the standard deviation, skew and the $(90^{th} - 10^{th})$ percentiles of the log pixel intensity. Figure 3-10 demonstrates how to interpret scatter plots. If the standard deviation of log luminance is used to estimate reflectance, the bounds of the quadratic fit are very loose. This is clearly seen at a standard deviation of 0.2, materials with reflectance in the range $(0.1, 0.9)$ may have a standard deviation of 0.2. Thus this statistic is not very useful for estimation. If we divide out data set into two categories Black (reflectance < 0.2) and White (> 0.6) i.e. ignore the green points on the scatter plot, then the statistic does a reasonable job at classification.

Figures 3-10(b) and 3-16(a) demonstrates the utility of statistics of luminance histograms in the form of ROC curves. Individual statistics by themselves achieve classification rates of $70 - 80\%$ on our data set of mean normalized materials. This is performance is certainly above chance and contrary to the Gelb effect, however it is not perfect. Figure 3-17 shows examples of misclassified images.

3.5 Statistics of Filtered Images

Next, like Dror et al [16], we examined the statistics of filtered images. Motivated by the filtering mechanisms in our visual system we consider the output of multi scale filtering on our images. Center surround and edge detection filters were used in a multi scale decomposition (Figure 3-12). Figure 3-13(a) and (b) show examples of such filtering. We examine the pixel intensity histograms for each subband image (the filtered image at each scale). The histograms of the subbands for light and dark material display typical differences (Figure 3-13(c)).

The histograms of dark materials tend to have higher standard deviation, heavier tails and for the case of center surround filtered images, are also skewed. The filters

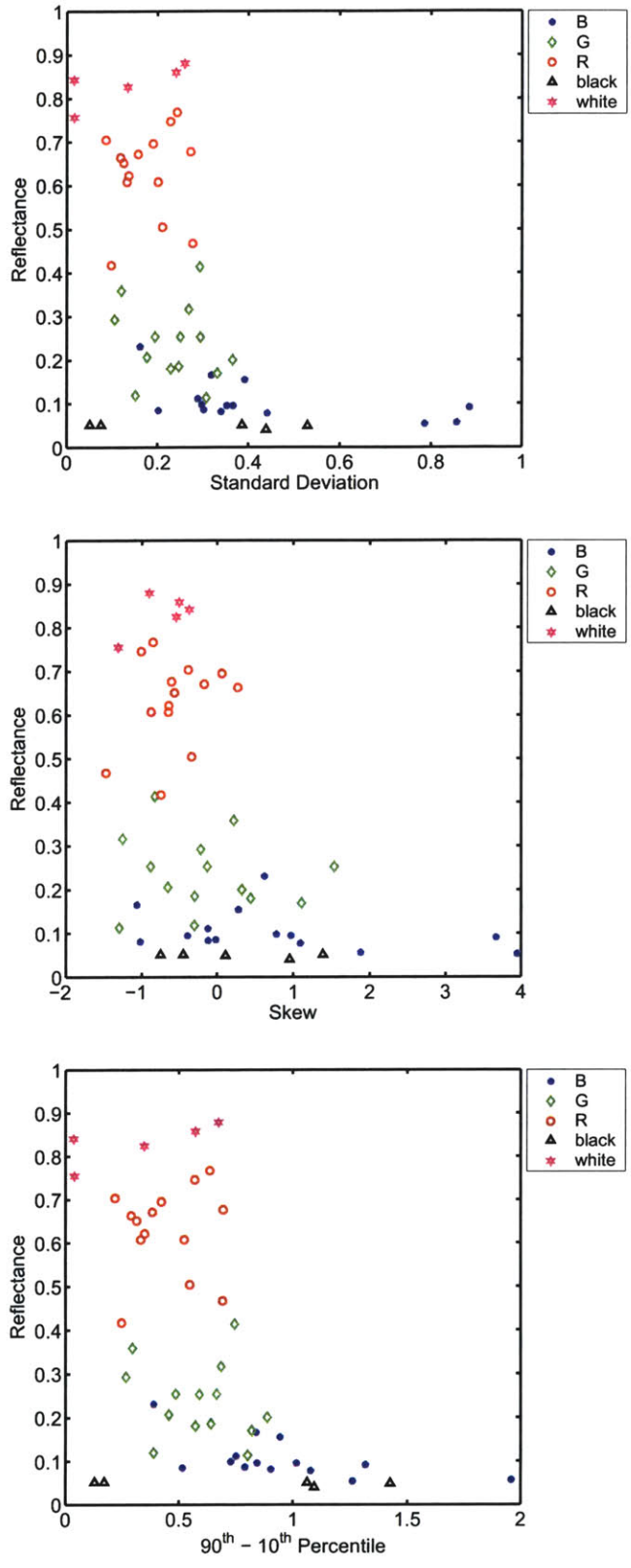
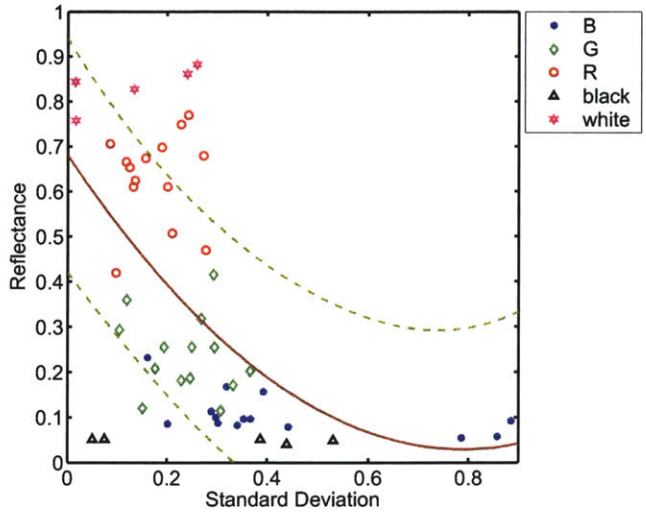
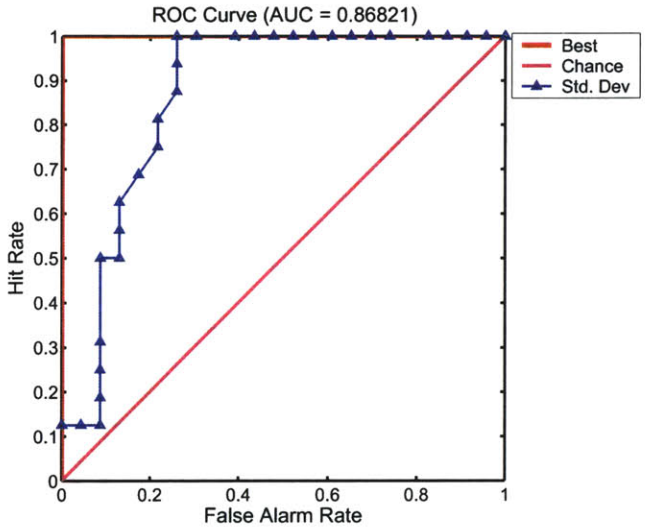


Figure 3-9: Ground Truth versus Statistics of the log(Luminance) Light 1



(a)



(b)

Figure 3-10: Interpreting Scatter Plots (a) Reflectance vs Standard Deviation of log(Luminance) for Light 1 - the best quadratic fit to the data is plotted with error bounds. If errors in fitting are independent and normally distributed with constant variance, the error bounds contain at least 50% of the data. (b) If we divide our image data into two sets BLACK (reflectance < 0.2) and WHITE (reflectance > 0.6), then standard deviation of log(Luminance) is a good feature for binary classification. The area under the ROC curve is 0.87.

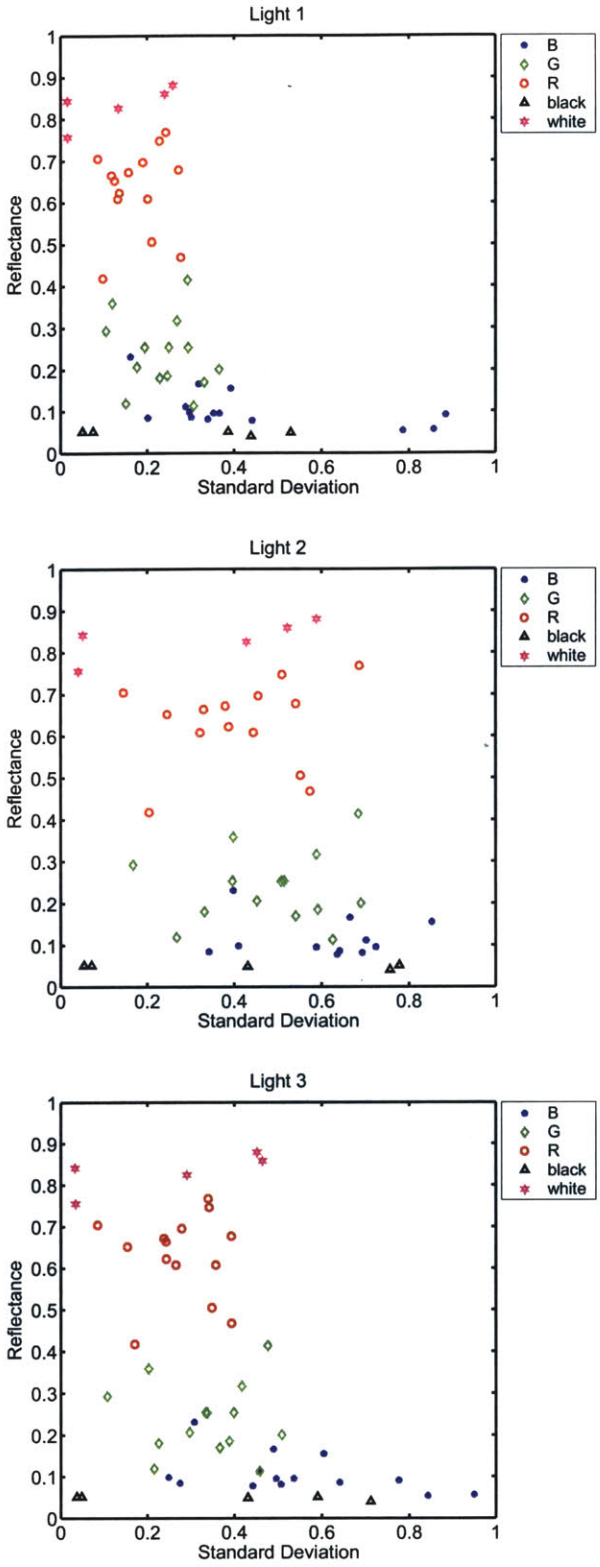


Figure 3-11: Reflectance vs Std. Deviation of log(Luminance) in Light 1 (top), Light 2 (middle) and Light 3(bottom)

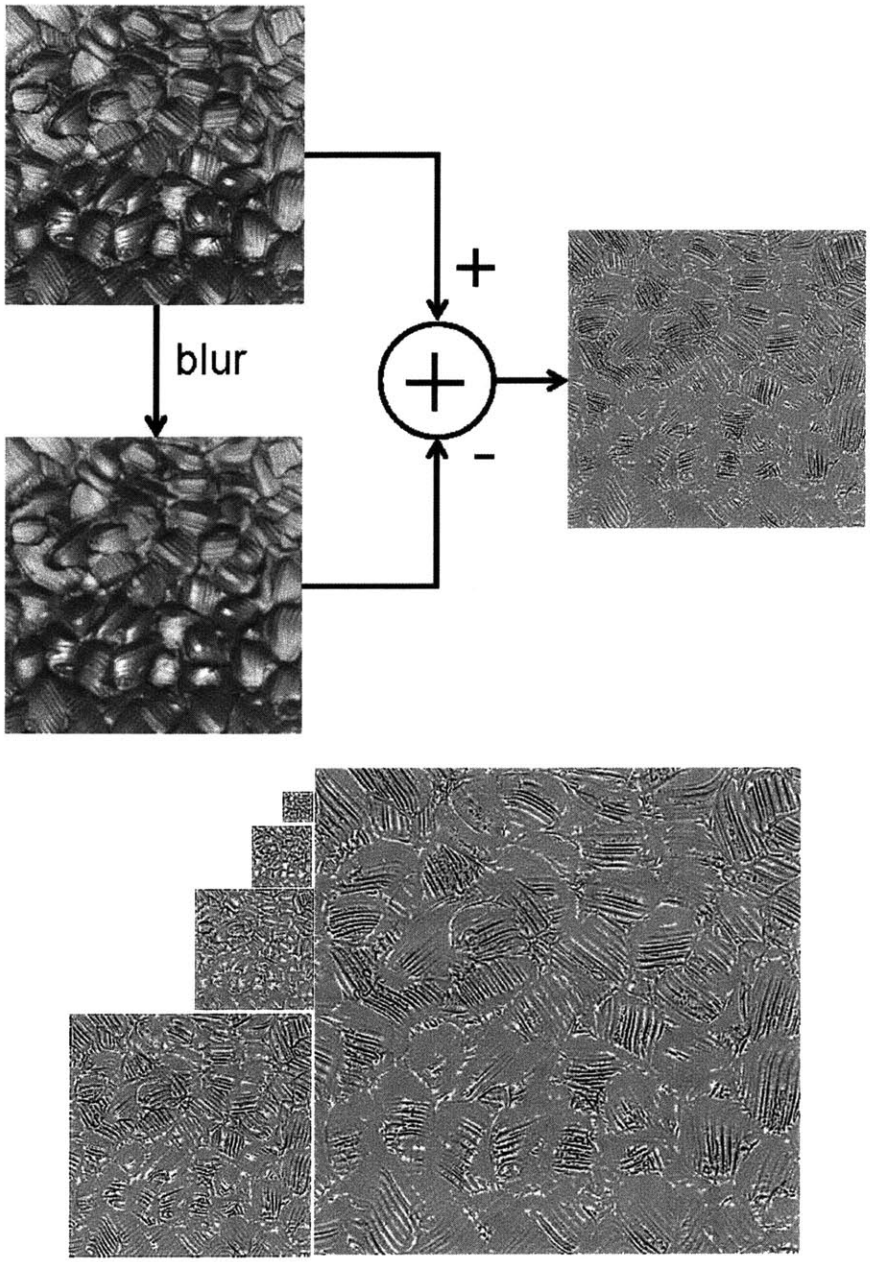


Figure 3-12: Center Surround Filtering: The mean normalized intensity image is compressed via a log transformation, then blurred by a gaussian filter ($\sigma = 1$, spatial support 5×5 pixels). The blurred image is subtracted from original to yield the center surround filtered image. This process is repeated at multiple scales.

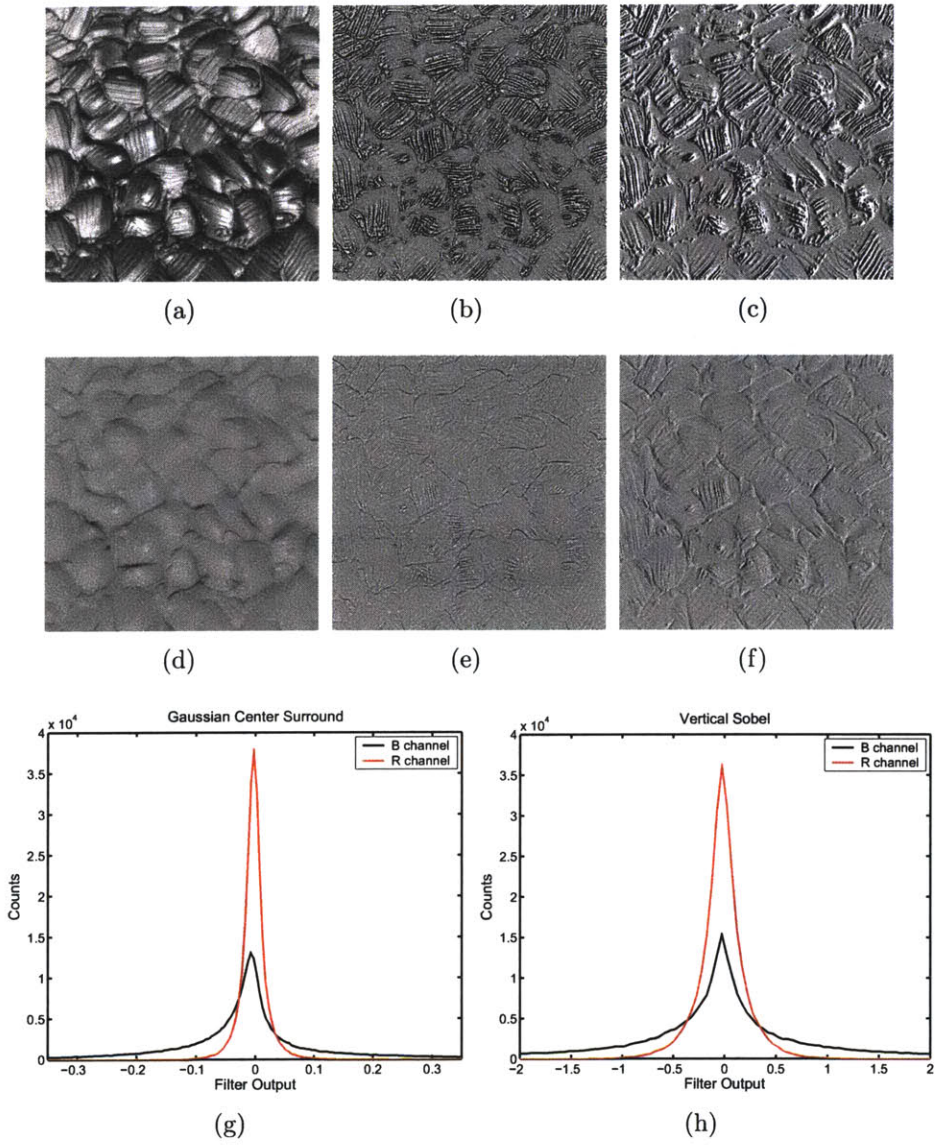


Figure 3-13: The statistics of filtered outputs are diagnostic of reflectance (a) Mean normalized B channel image (b) Gaussian center surround filtered B (c) Vertical Sobel filtered B (d) Mean normalized R channel image (e) Gaussian center surround filtered R (f) Vertical Sobel filtered R (g) Pixel Histogram of (b) and (e) (h) Pixel Histogram of (c) and (f). Typically the histograms for the B channel (black) have higher standard deviation, longer tails and for the center surround case, are also skewed.

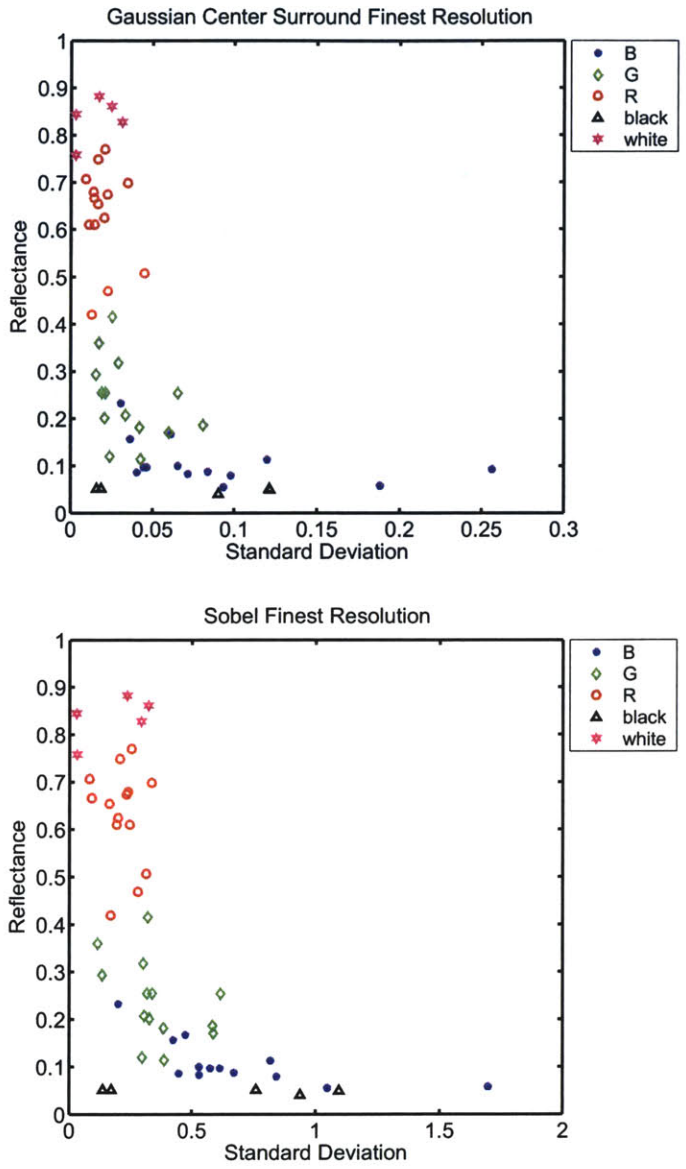


Figure 3-14: Reflectance vs Std. Deviation of filter output in Light 1 (top) Gaussian Center Surround (bottom) Vertical Sobel. For both plots, the filter was applied to the log(mean normalized pixel intensity) image at full resolution.

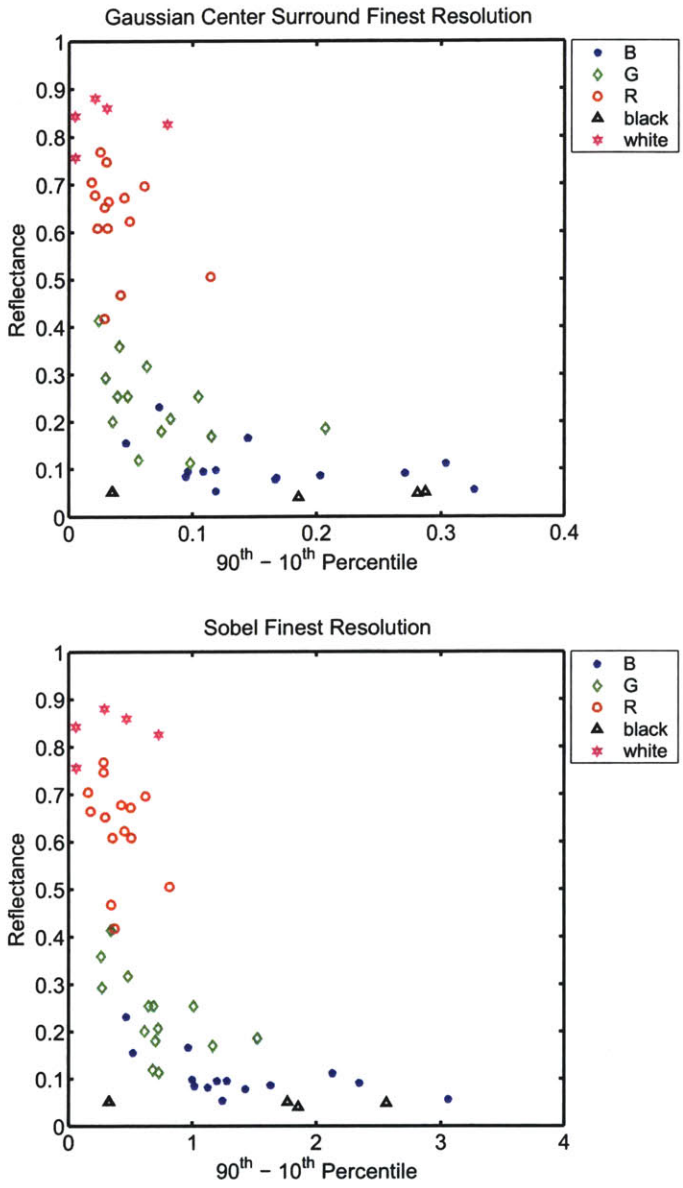
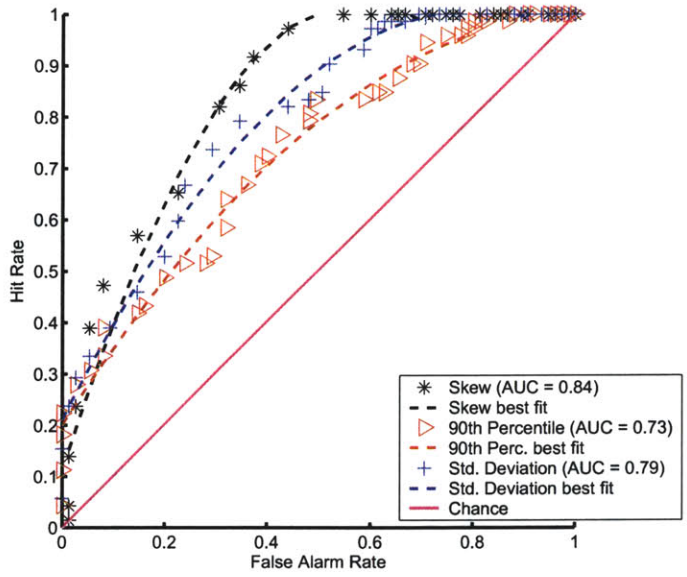
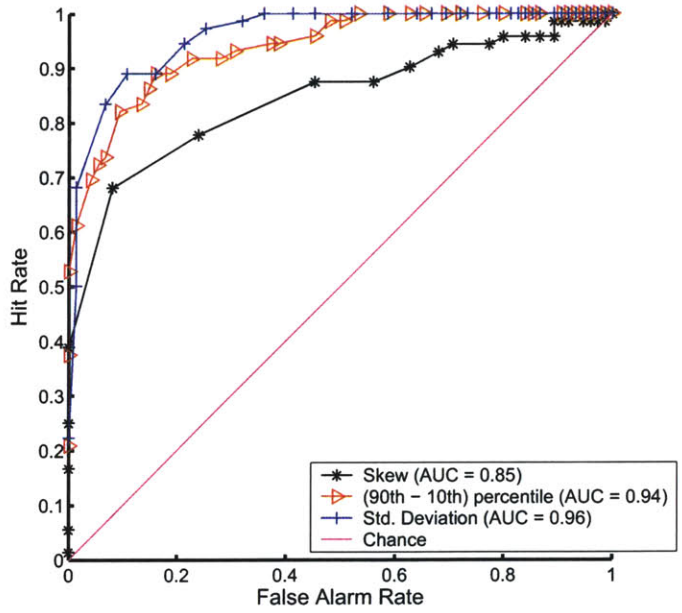


Figure 3-15: Reflectance vs ($90^{th} - 10^{th}$) Percentile of filter output in Light 1 (top) Gaussian Center Surround (bottom) Vertical Sobel. For both plots, the filter was applied to the log(mean normalized pixel intensity) image at full resolution.

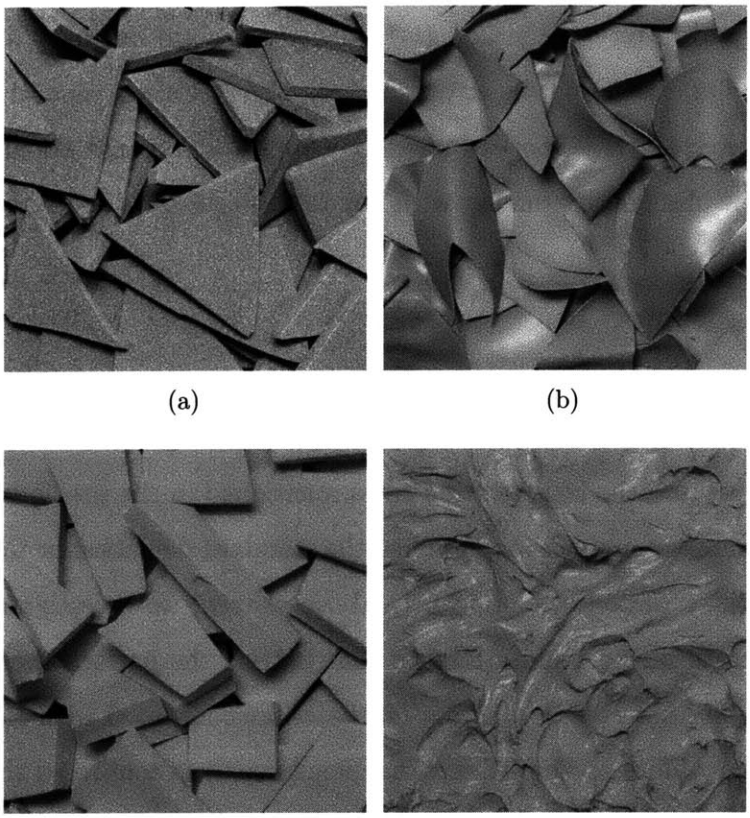


(a)



(b)

Figure 3-16: ROC Curves for (a) statistics of the intensity histogram and (b) statistics of the histogram of filtered intensities. For this plot a gaussian center surround filter was used at the finest resolution.



(a) (b)

(c) (d)

Figure 3-17: Examples of misclassified materials

pick up on the deep shadows, higher local contrast and brighter specularities of dark materials thus resulting in these typical characteristics.

The moment and percentile statistics of intensity histograms of subband images capture these characteristics and ROC analysis reveals their significance for reflectance classification (Figure 3-16(b)). Individual statistics by themselves achieve classification rates of greater than 85% on our data set. The performance is way above chance and is very encouraging. However there are mistakes in classification (Figure 3-17). On examining the errors made by the statistics, we find that for many images in Figure 3-17 it is hard to make perceptual judgments about their albedo. We explore this observation further with psychophysical tests in Chapter 5.

3.6 Interdependence of the statistics

The ROC curves and the classification rates mentioned in the previous sections refer to the performance of a *single* statistic like the standard deviation or the skew. Given that each statistic performs significantly above chance, we investigated if we could combine the statistics to build a feature that betters the performance of individual statistics. We find however that all our statistics are correlated with each other. Mutual information values and chi-square independence tests confirmed this observation (Figure 3-18, Table 3.5).

This dependence may be because the chosen statistics are inter-dependent in all natural images. To test this hypothesis we measured the same statistics on some images of natural scenes. We find that the statistics were not correlated, had lower values of mutual information and passed the chi-square independence test. These results negate our hypothesis, so we must search for other causes for the dependence.

The reason for the dependence in our statistics is most likely an artifact of our

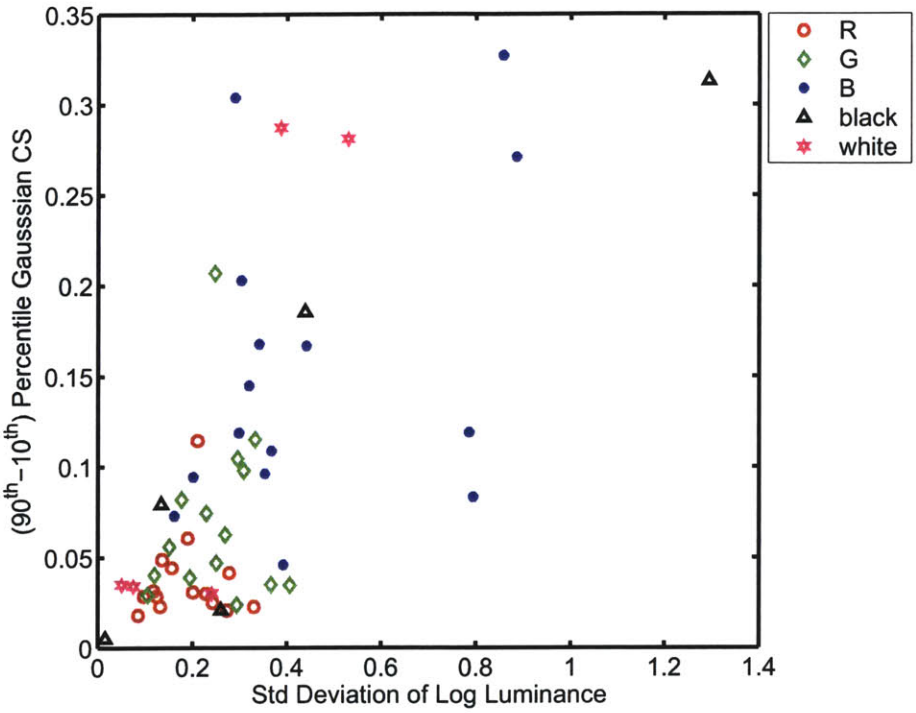


Figure 3-18: Standard Deviation of log(Luminance) vs $(90^{th} - 10^{th})$ Percentile of gaussian center surround filtered image (finest resolution) for Light 1. Combining two statistics is not significantly better than using a single statistic to predict the reflectance of a material. For the plot above, $r = 0.6732$ ($p < 0.05$).

materials data set. The materials with lower reflectance have higher local contrast as well as high frequency structure. Therefore, the statistics that measure local contrast (standard deviation of intensity histograms or of histograms of center surround filter outputs) are dependent on the statistics that measure energy in high frequencies ($(90^{th} - 10^{th})$ percentile of histograms of filter outputs at finest resolution).

To conclude, our statistics are not independent of each other, yet from Table 3.5 we observe that they are not completely correlated. It would be interesting to explore how much information they share and the smallest subset of statistics that optimizes the information shared by its members. In practice, we find that optimal performance at reflectance classification or estimation is achieved using any three of the statistics listed in Table 3.5.

Feature Index	Feature Index	Correlation Coefficient		Mutual Information (bits)	Chi Square Independence $c = 26.2962$	
		r	p		T	$T < c$
1	1	1.0000	0	2.4665	672.0060	0
	2	0.4132	0	0.2461	55.8360	0
	3	0.8869	0	1.2045	209.6778	0
	4	0.5859	0	0.3955	70.5536	0
	5	0.3779	0	0.2084	36.1627	0
	6	0.5883	0	0.3251	82.4198	0
2	2	1.0000	0	1.8467	672.0060	0
	3	0.7085	0	0.4590	142.8867	0
	4	0.3658	0	0.2235	42.3649	0
	5	0.7779	0	0.5200	123.0237	0
	6	0.6904	0	0.4014	114.7801	0
3	3	1.0000	0	2.0186	672.0060	0
	4	0.5861	0	0.3355	64.2640	0
	5	0.5882	0	0.2845	54.9790	0
	6	0.8021	0	0.4582	165.9946	0
4	4	1.0000	0	2.0184	672.0060	0
	5	0.2879	2×10^{-4}	0.2440	19.3077	1
	6	0.8065	0	0.7424	151.9679	0
5	5	1.0000	0	1.5595	672.0060	0
	6	0.5165	0	0.3145	46.1386	0
6	6	1.0000	0	1.3909	672.0060	0

Table 3.5: **Dependence amongst statistics** Feature index is 1 = ($90^{th} - 10^{th}$) percentile of log(Luminance), 2 = skew of log(Luminance), 3 = standard deviation of log(Luminance) and similarly 4 = ($90^{th} - 10^{th}$) percentile, 5 = skew and 6 = standard deviation of gaussian center surround filtered image (finest resolution). The correlation coefficient for every pair of features is significant ($p < 0.05$) and quite high. The value of mutual information for any pair cannot be ignored. Finally, except for feature pair $(4, 5)$ all other pairs fail the Chi Square test of independence.

Chapter 4

Synthesizing material appearance

In the previous chapter we observed that image statistics like moments and percentiles of image intensities or of filtered image intensities, are useful for predicting the reflectance of a material. Changing the material (and therefore its physical reflectance) leads to changes in image statistics of its photographs. It would be nice to know if the reverse is true i.e. do changes in image statistics correspond to a change in the perceived reflectance of the material?

It is possible to manipulate an individual statistic (moment or percentile) of an image. However, as all our statistics are interdependent (Section 3.6) it is hard, if not impossible, to manipulate them *independently* of each other, in a way that the resulting image looks like a photograph of a real world material. This sort of image manipulation is related to the problem of texture synthesis-by-analysis. Texture analysis is the problem of identifying a texture metric. A texture metric operates on two images and defines the distance between them in the perceptual texture space. A texture metric could be a set of image features or statistics. Texture synthesis is the problem of generating a new sample of a texture, given an exemplar. The power of a texture metric is revealed when performing texture synthesis. The synthesized texture image and original exemplar image, must satisfy two conditions - first, the texture metric must recognize both images as belonging to the same texture class and second, to humans the images should look like they correspond to the same texture

class. Texture synthesis-by-analysis proceeds by manipulating an image so that the *distance* (as defined by the texture metric) between the manipulated image and the exemplar texture is minimized. Our situation is analogous, we have a set of image statistics that are diagnostic of reflectance and given an example material image, we want to synthesize a new image of a surface that has the same image statistics and is perceived to be of the same reflectance as the example.

There is a large body of work in texture analysis and synthesis [25, 60, 18, 41]. The work that influences us the most is that of Heeger & Bergen [25].

4.1 Previous Work

The Heeger-Bergen algorithm [25] takes as input an example "target" texture. The goal of the algorithm is to synthesize an image that visually appears to be a new sample of the target texture. The authors are influenced by recent work in human texture perception. Theories of texture discrimination state that two textures appear similar when they produce similar distributions of responses in a bank of linear, spatial frequency selective filters. Therefore, the texture metric in the Heeger-Bergen algorithm is defined as the distance between intensity histograms and histograms of filter bank outputs.

The synthesized image is initialized to a random noise texture. The algorithm proceeds by iteratively matching the intensity histogram and the subband histograms of a steerable pyramid decomposition of the noise texture to those of the target texture. The algorithm converges in about five iterations though there is no formal proof for convergence. The pseudo code is reproduced from the original paper in Figure 4-1. Histogram matching is a generalization of the histogram equalization procedure. The source histogram is matched to the target histogram by constructing the cumulative distribution function of the source image and the inverse cumulative distribution

function of the target image. The pseudo code for the histogram matching procedure is shown in Figure 4-2.

```
Match-texture(noise,texture)
    Match-Histogram (noise,texture)
    analysis-pyr = Make-Pyramid (texture)
    Loop for several iterations do
        synthesis-pyr = Make-Pyramid (noise)
        Loop for a-band in subbands of analysis-pyr
            for s-band in subbands of synthesis-pyr
            do
            Match-Histogram (s-band,a-band)
    noise = Collapse-Pyramid (synthesis-pyr)                    10
    Match-Histogram (noise,texture)
```

Figure 4-1: Pseudo Code for Heeger-Bergen reproduced from [25]

```
Match-histogram (im1,im2)
    im1-cdf = Make-cdf(im1)
    im2-cdf = Make-cdf(im2)
    inv-im2-cdf = Make-inverse-lookup-table(im2-cdf)
    Loop for each pixel do
        im1[pixel] = Lookup(inv-im2-cdf,Lookup(im1-cdf,im1[pixel]))
```

Figure 4-2: Pseudo Code for Match-Histogram procedure in Figure 4-1 (reproduced from [25])

The Heeger-Bergen texture synthesis work is considered seminal because it combines ideas from texture analysis, statistics and psychophysics in a remarkably simple algorithm. The algorithm produces good results for stochastic textures but fails for structured textures. The success or failure of the Heeger-Bergen algorithm may be attributed to two factors - the choice of texture metric and the search procedure for obtaining the final synthesized texture. We find that the search procedure of the Heeger-Bergen algorithm (iterative histogram matching) succeeds for most images (Figure 4-6). Therefore, the failures of the algorithm may be attributed to the choice of texture metric. The failures illustrate that distance between intensity histograms

and subband histograms is an inadequate texture metric.

Zhu et al [60] and Portilla and Simoncelli [41] expand the Heeger-Bergen texture metric to include non-linear filters and joint distributions of subband coefficients respectively. The search procedures for these algorithms are mathematically involved but both achieve significantly better results than Heeger-Bergen.

There is a whole class of non-parametric texture synthesis techniques [18, 17] that achieve impressive results. However non-parametric techniques are not useful for us because the texture metric in these techniques is not explicitly defined. Therefore it is hard to cast our problem of synthesizing material appearance in their framework.

4.2 Luminance Histogram Equalization

In Chapter 3 we observed that moment and percentile statistics of the luminance (or log luminance) histogram are useful features for predicting reflectance. To synthesize a new material image given target and source images, a first thought is to equalize the luminance histograms. Figures 4-3 and 4-4 show examples of such a procedure. Given a source image (say R channel of an orange material) and a target image (say B channel of same material), if the statistics of the luminance histogram are what determine the perceived reflectance, then matching the histogram of the source to the target, should produce an image perceptually identical to the target. However, this is not the case.

In Figure 4-4 we observe that B2R and R2B images are still perceptually distinguishable from target R and B images respectively. Therefore, statistics of the luminance histogram are sufficient reflectance descriptors for some materials, but not all.

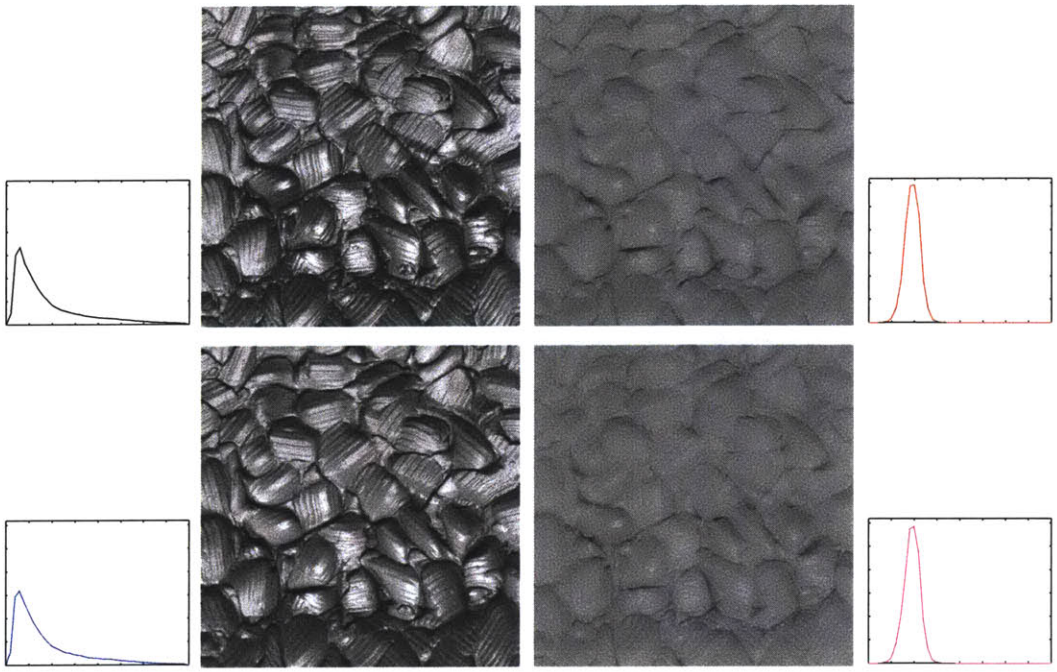


Figure 4-3: **More successful example of histogram equalization** (Top left) Original B channel image and its luminance histogram in black to its left (Top Right) Original R channel image and its luminance histogram in red to its right (Bottom Left) R2B : Histogram of R image is forced to be the same as B histogram. The histogram of the result R2B is indicated in blue to the left (Bottom Right) B2R : Histogram of B image is forced to be the same as R histogram. The histogram of the result B2R is indicated in magenta to the left. The histogram of B and R2B are very similar. The Chi Square distance between them is $\chi^2(h_B, h_{R2B}) = 0$ $(p = 0)$. Histogram of R and B2R are also close, $\chi^2(h_R, h_{B2R}) = 0.0026$ $(p = 0)$.

Figure 4-4: **Less successful example of histogram equalization** (Top left) Original B channel image and its luminance histogram in black to its left (Top Right) Original R channel image and its luminance histogram in red to its right (Bottom Left) R2B : Histogram of R image is forced to be the same as B histogram. The histogram of the result R2B is indicated in blue to the left (Bottom Right) B2R : Histogram of B image is forced to be the same as R histogram. The histogram of the result B2R is indicated in magenta to the left. The histogram of B and R2B are very similar. The Chi Square distance between them is $\chi^2(h_B, h_{R2B}) = 0.0062$ $(p = 0)$. Histogram of R and B2R are also close, $\chi^2(h_R, h_{B2R}) = 0.0126$ $(p = 0)$.

4.3 Heeger-Bergen applied to material images

As the statistics of the luminance histograms do not capture everything, the next thought is to use the Heeger-Bergen algorithm directly. In addition to the luminance histogram, Heeger-Bergen enforces the histograms of subbands of the source image to match the histograms of subbands of the target image. As discussed in Section 4.1, this is accomplished by an iterative procedure that converges, in practice, in a few iterations to the desired target image. A subsampled steerable image pyramid representation [47] is used to analyze the frequency content of the image.

We know that the statistics of histograms of filtered outputs are diagnostic of reflectance. By enforcing the source and target image to have to the same luminance histogram and the same subband histograms, we expect the synthesized image to be closer in reflectance to the target than source image. In fact, one would expect the synthesized image to be perceptually closer than what plain luminance histogram equalization would get us. Figure 4-5 shows the result of running Heeger-Bergen on Material 1. Both the results B2R and R2B look pretty bad. While these results are not atypical of the Heeger-Bergen algorithm, these images do not suffice for our purposes. They do not look like natural images and it is unfair to ask observers to judge the reflectance of such images.

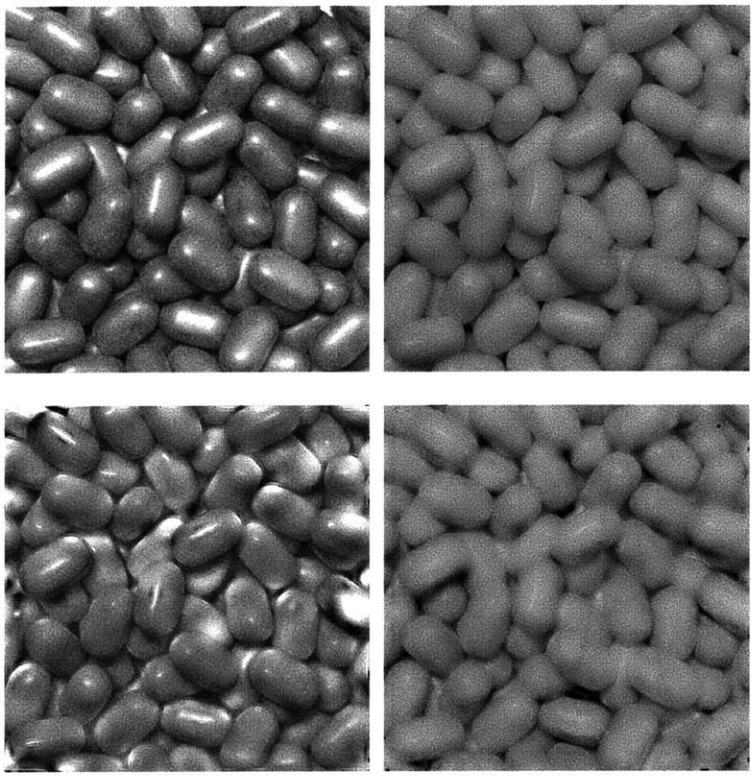
Figure 4-6 illustrates that the failure in Figure 4-5 R2B image is not because the algorithm fails to match the subband histograms or the luminance histogram, rather there is a problem in our formulation of using Heeger-Bergen in its original form on material images. Figure 4-7 is an example of moderate success with direct Heeger-Bergen on a different material.

The artifacts observed in Figures 4-5 and 4-7 are because histogram matching applies a pointwise non-linear gain to the values in a subband which may lead to local distortions (refer 4.4 for details). Moreover, when histograms at each scale are ma-

nipulated independently of each other, there is no guarantee that the contributions at any point in the synthesized image from the different subbands agree with each other.

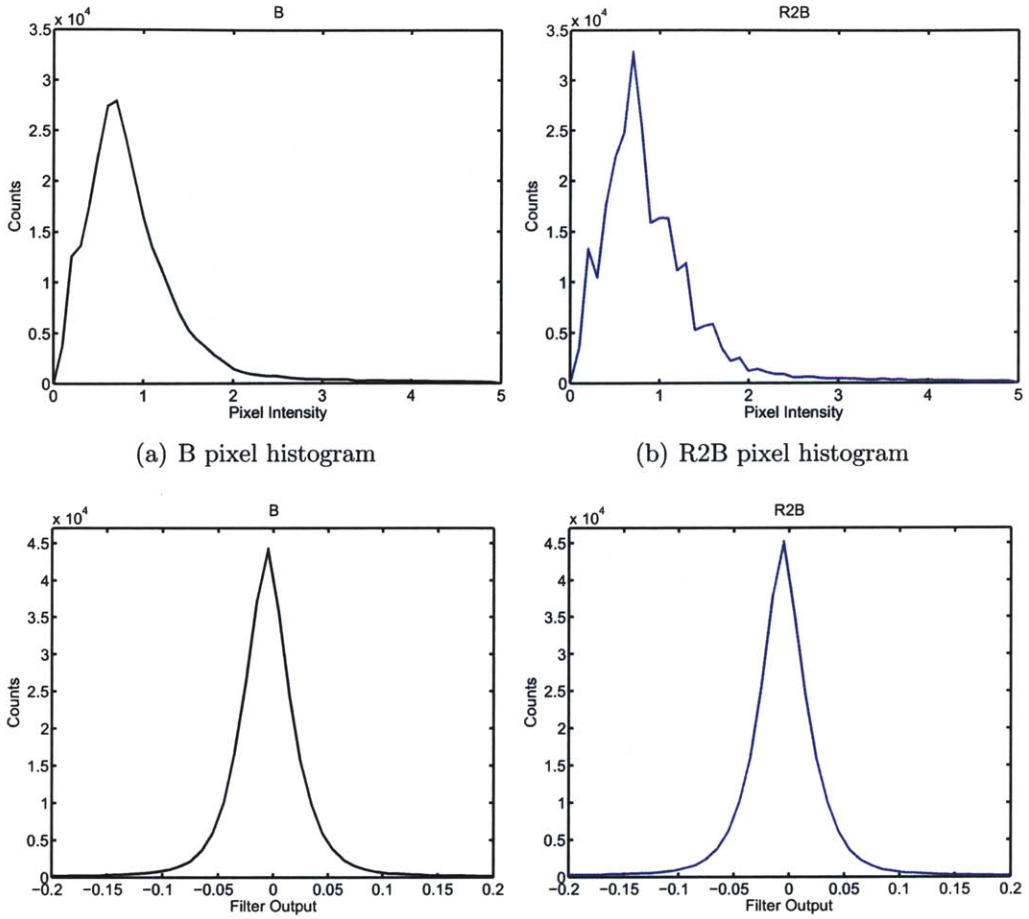
Solutions for reducing the observed artifacts include using a rotationally symmetric transform, such as a Laplacian pyramid and avoiding subsampling in the pyramid. A rotationally symmetric transform has no orientation dependent subbands hence the need for agreement between orientation subbands at each scale is eliminated. Image pyramids [7, 46, 22] are downsampled by a factor of two at each scale for computational efficiency. By avoiding subsampling and introducing more redundancy in the representation, we can avoid some aliasing artifacts.

Figures 4-8 and 4-9 show results of using a subsampled Laplacian pyramid instead of subsampled steerable pyramid in the Heeger-Bergen procedure. The results are markedly better. In the next section we propose an improvement that produces further improvements. It is important to note that we use a simple non-oriented pyramid (Laplacian pyramid) and simple image statistics (moments and percentiles) in the Heeger-Bergen framework because we do not synthesize an image from scratch (e.g. starting with random noise). We assume an image of a material is provided to us and we want to preserve the structure of the image and only manipulate the appearance of the material.



(c) (d)

Figure 4-5: **An Example of Failure with Heeger Bergen** Original Heeger Bergen with subsampled steerable pyramid (Scales = 4, Orientations = 4)(a) B image (b) R image (c) R2B : R image is the initial texture and B image is the target texture (d) B2R : B image is the initial texture and R image is the target texture.



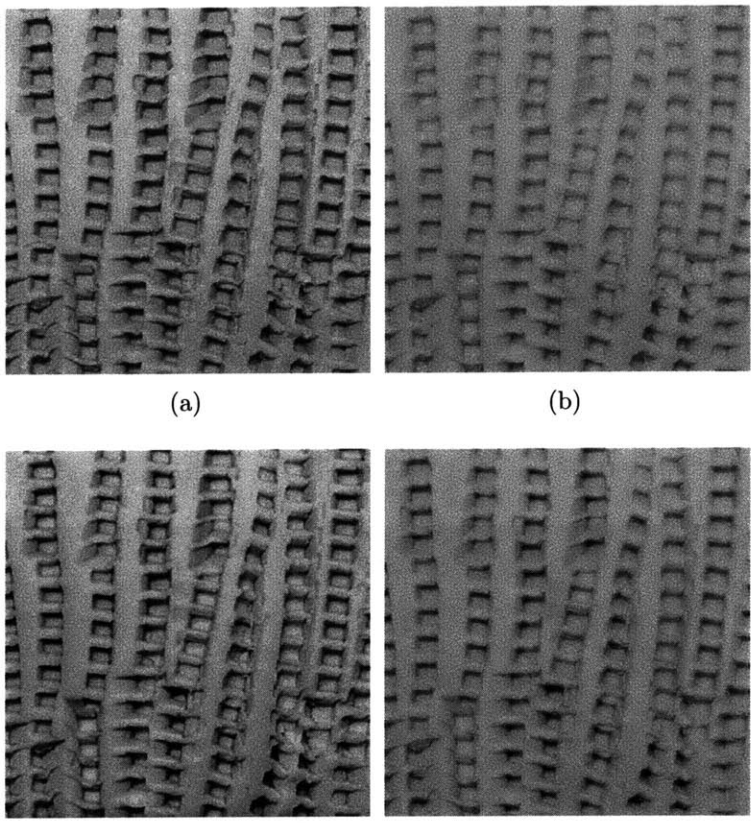
(a) B pixel histogram

(b) R2B pixel histogram

(c) B subband histogram

(d) R2B subband histogram

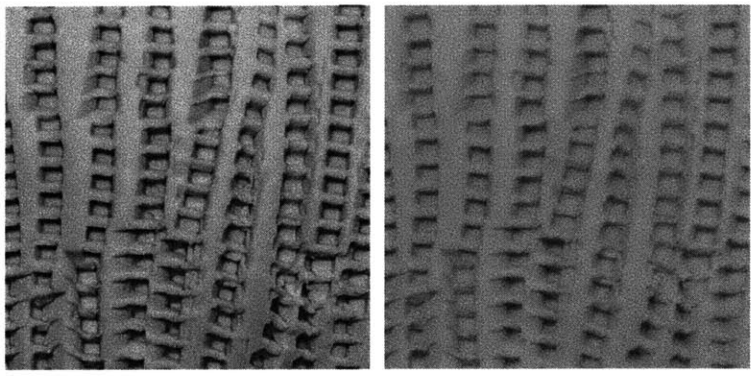
Figure 4-6: **Direct Heeger Bergen search succeeds** (a) Pixel histogram of B image (Figure 4-5a) (b) Pixel histogram of R2B image (Figure 4-5c)(c) Pixel histogram of an oriented subband of B image (finest scale, diagonal orientation) (d) Pixel histogram of an oriented subband of R2B image (finest scale, diagonal orientation). The pixel histograms of B and R2B image are well matched $\chi^2(h_B, h_{R2B}) = 0.0117$ $(p = 0)$ and so are the subband histograms $\chi^2(h'_B, h'_{R2B}) = 0.0048$ $(p = 0)$.



(a) (b)

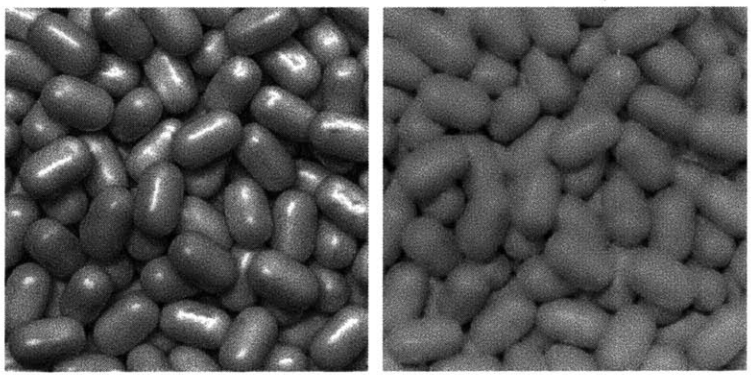
(c) (d)

Figure 4-7: **Example of success with Heeger Bergen** Original Heeger Bergen with subsampled steerable pyramid (Scales = 4, Orientations = 4)(a) B image (b) R image (c) R2B : R image is the initial texture and B image is the target texture (d) B2R : B image is the initial texture and R image is the target texture.



(a) (b)

Figure 4-8: **Heeger Bergen with subsampled Laplacian pyramid** (Scales = 4)(a) R2B : R image is the initial texture and B image is the target texture (b) B2R : B image is the initial texture and R image is the target texture.



(a) (b)

Figure 4-9: **Another example of Heeger Bergen with subsampled Laplacian pyramid** (Scales = 4)(a) R2B : R image is the initial texture and B image is the target texture (b) B2R : B image is the initial texture and R image is the target texture.

4.4 Activity Map based Heeger-Bergen

The histogram matching procedure in the Heeger-Bergen algorithm (Figure 4-2) involves applying a non-linear *pointwise* gain, i.e. each pixel value in an image is mapped to a new value, independent of other pixels. A pointwise gain is not desirable because if the value of a pixel is manipulated independently of its neighbors local distortions can occur. We propose the following solution - instead of matching histograms of the source and target subbands directly, we will modify the source histograms by manipulating activity maps. Let us define an activity map as the result obtained by taking the absolute value of a subband and then blurring it with a gaussian kernel (Figure 4-10).

The combination of absolute value and blurring transforms the subband image into a local energy map. If we match the histograms of the activity maps of the target and source images, then a pointwise gain is applied to the source activity map. As the activity map may be thought of as a local energy map, a pointwise gain on the source activity map is effectively a *local* gain on the original subband. Let the original source activity map be A_{orig} and the histogram matched source activity map be $A_{modified}$. Then, the gainmap G is calculated as

$$G = \frac{A_{modified}}{A_{orig}} \tag{4.1}$$

G is multiplied to the original source subband to obtain the modified subband. Therefore, matching the histograms of the activity maps, allows us to apply a *spatially local* gain which results in fewer image artifacts and smoother looking pictures. The local gain modifies the value at a pixel depending on the values of its neighbors, therefore locally the distortions introduced by histogram matching are reduced.

Figures 4-11 and 4-12 show one iteration of activity map based Heeger-Bergen using an oversampled Laplacian pyramid. Figure 4-13 plots the Kulback-Leibler dis-

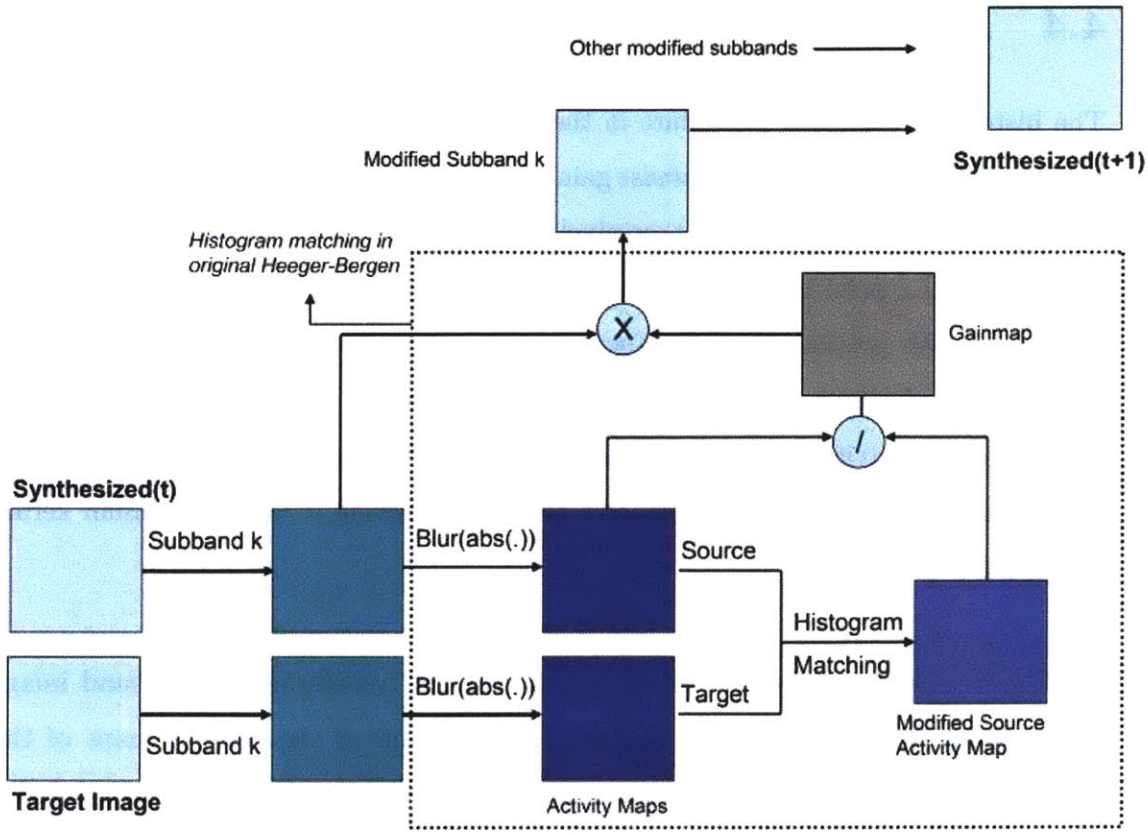


Figure 4-10: Activity Map Based Heeger-Bergen

tance between the histograms of the synthesized and target images as a function of the number of iterations. The results of this approach are compared with the approaches described earlier in Figures 4-14, 4-15. As with the regular Heeger-Bergen, the choice of image pyramids can be varied and subsampling avoided to obtain more pleasing results. Figures 4-16 through 4-20 show additional results.

Note that in Figures 4-16 through 4-20, we *map* the R channel of a material to its B channel or vice versa. If we apply our technique to the case where the source and the target images are of different materials, we get mixed results. (Figure 4-21).

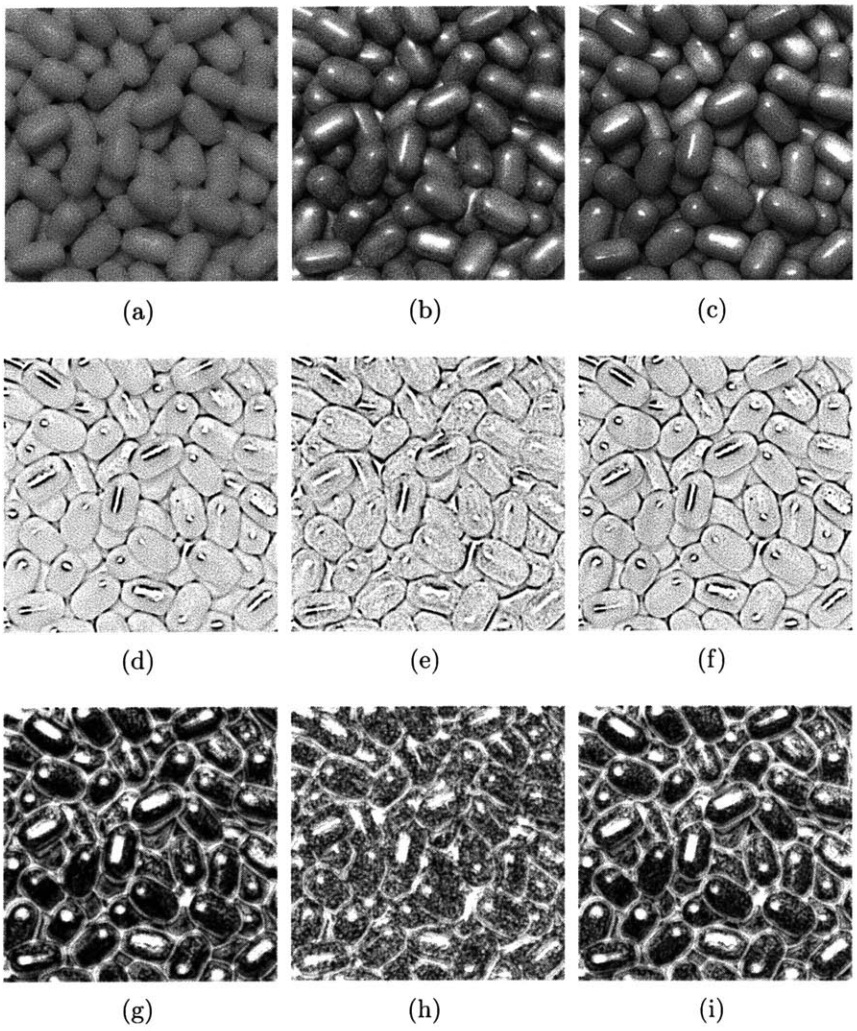
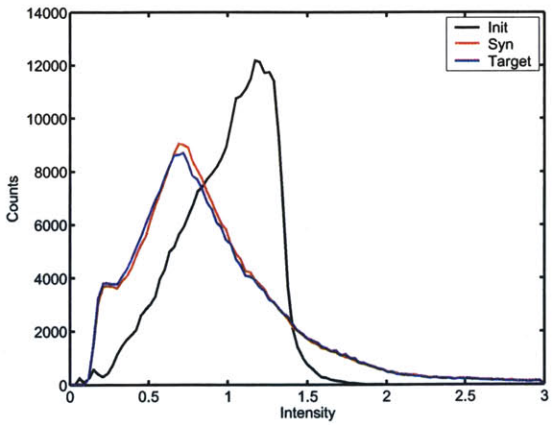
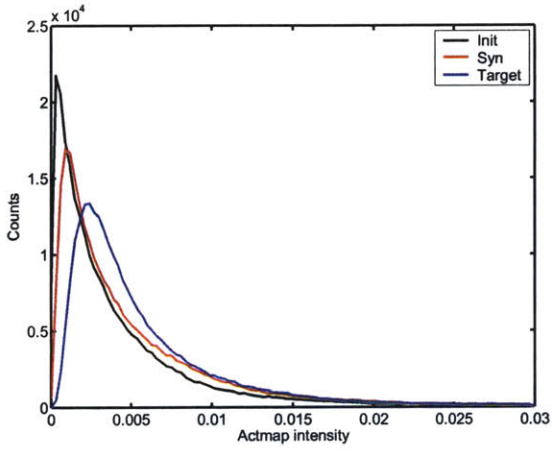


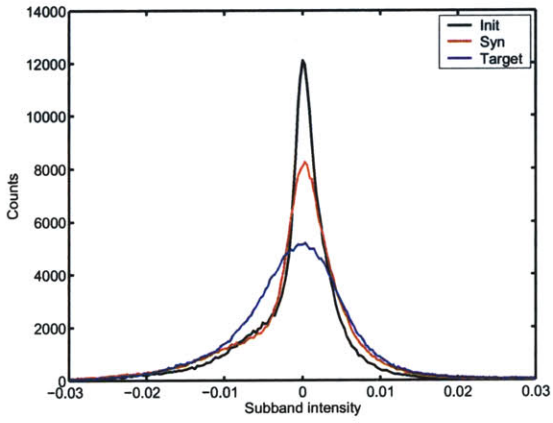
Figure 4-11: One Iteration of Activity Map based Heeger-Bergen (a) Initial R Image (b) Target B Image (c) Synthesized Image after one iteration. All images in (d)-(i) are derived from the third finest subband of the oversampled Laplacian image pyramid. (d) Subband of R (e) Subband of B (f) Subband of synthesized image (g) Activity map of R (h) Activity map of B (i) Activity map of synthesized image



(a)

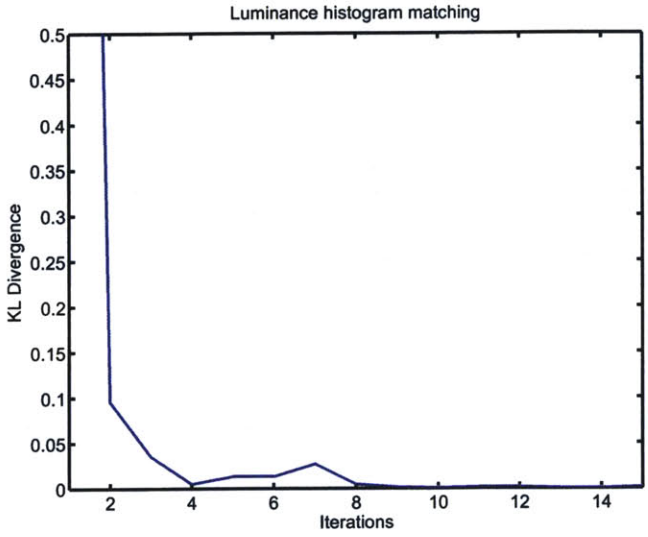


(b)

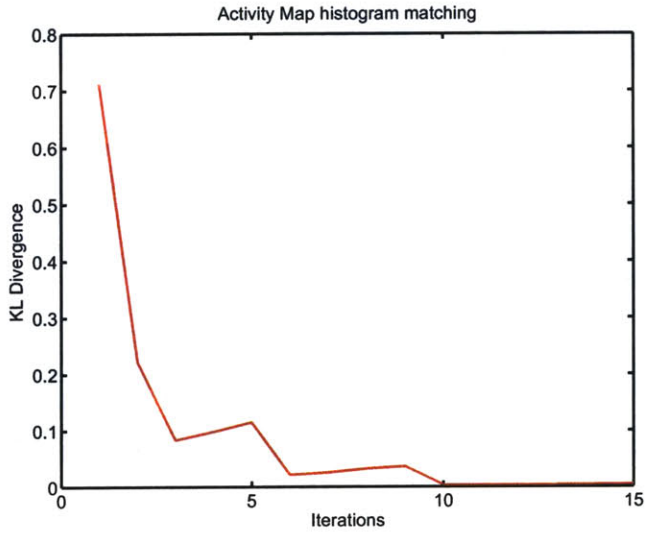


(c)

Figure 4-12: One Iteration of Heeger-Bergen (a) Luminance histograms (b) Activity map histograms (c) Subband histograms



(a)



(b)

Figure 4-13: KL Distance vs Iterations (a) KL Distance between luminance histograms of synthesized and target (b) KL Distance between Activity map histograms at third finest level

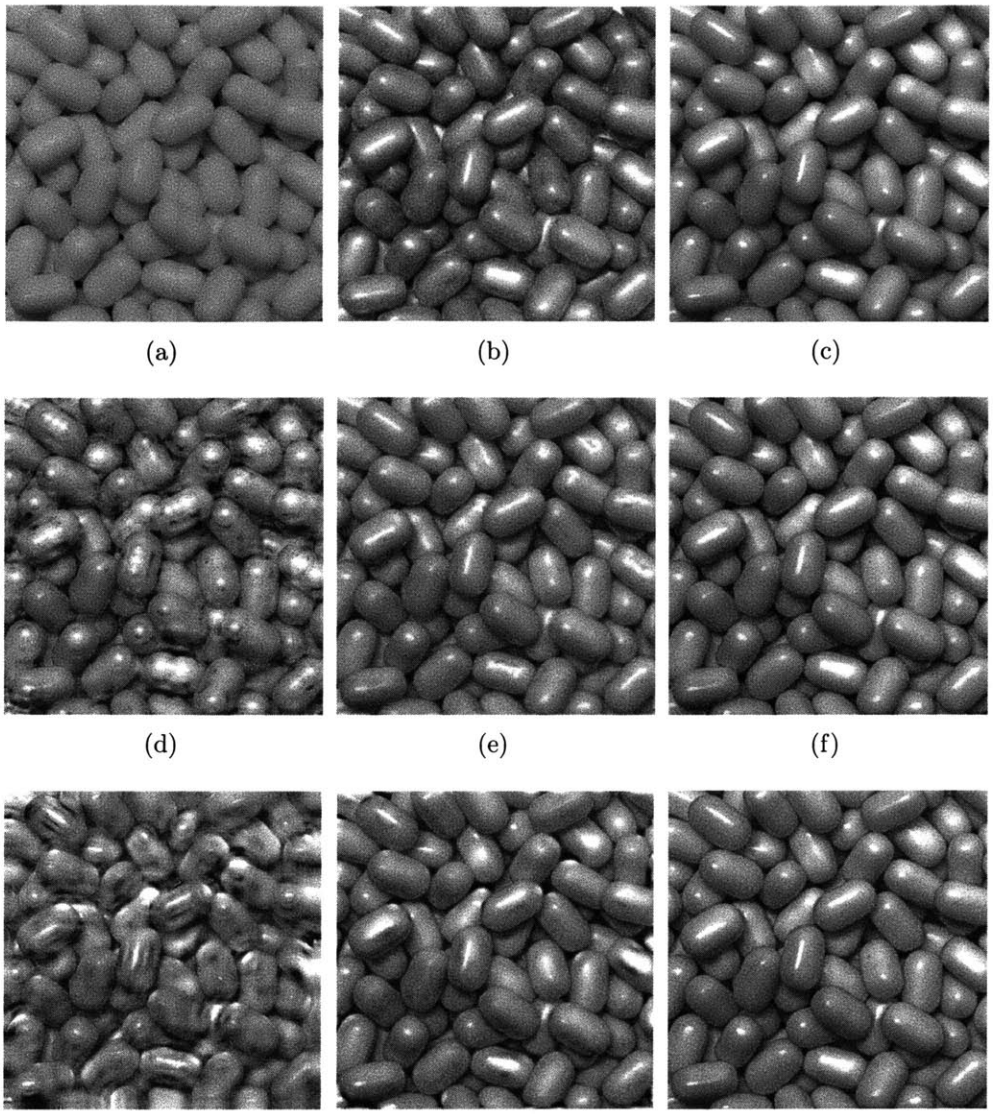


Figure 4-14: **Comparison of methods** (a) Initial R image (b) Target B image (c) Luminance histogram equalization (d) Heeger-Bergen steerable subsampled (e) Heeger-Bergen Laplacian subsampled (f) Heeger-Bergen Laplacian oversampled (g) Activity map based Heeger-Bergen Steerable subsampled (h) Activity map Laplacian subsampled (i) Activity map Laplacian oversampled

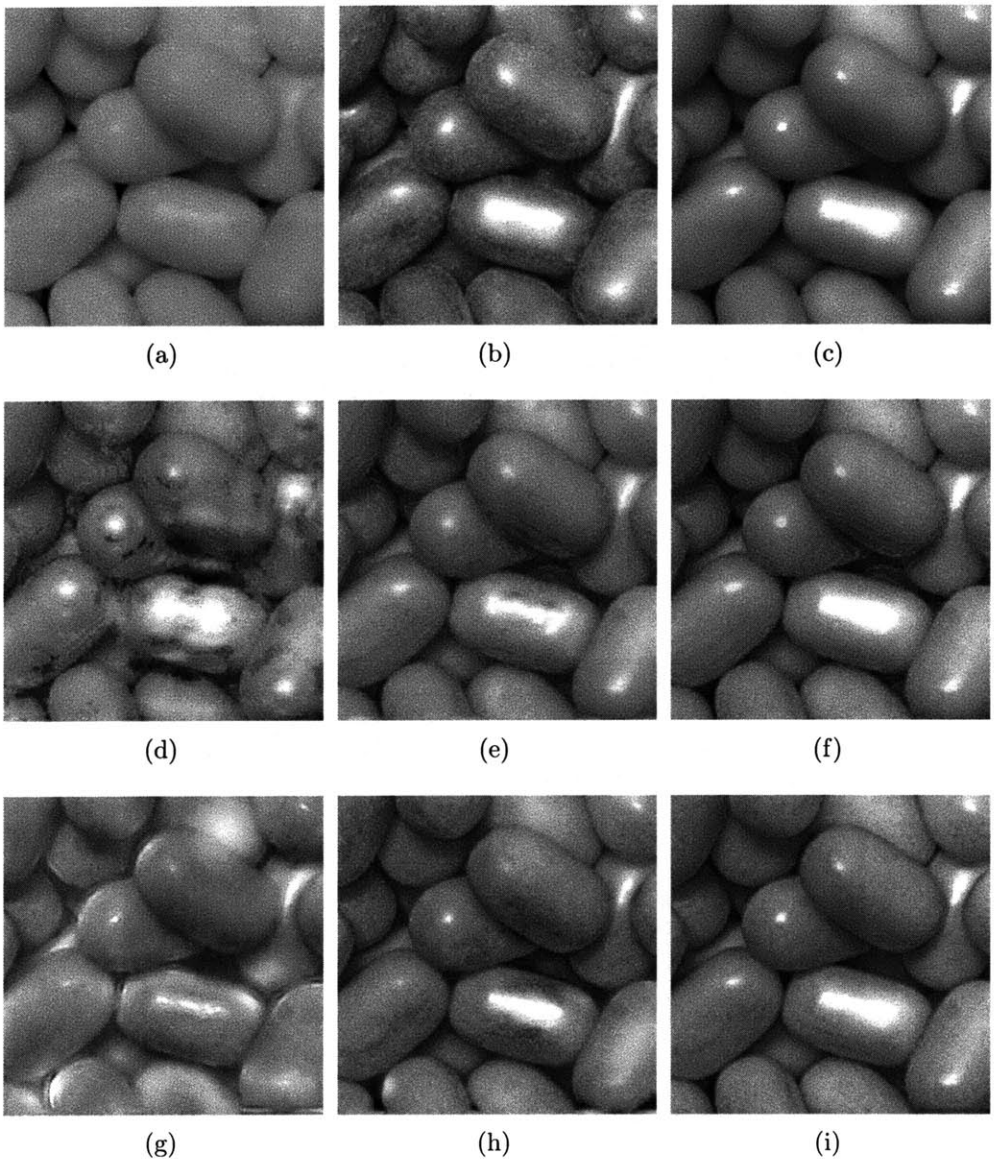


Figure 4-15: **Comparison of methods (Blowup)** (a) Initial R image (b) Target B image (c) Luminance histogram equalization (d) Heeger-Bergen steerable subsampled (e) Heeger-Bergen Laplacian subsampled (f) Heeger-Bergen Laplacian oversampled (g) Activity map based Heeger-Bergen Steerable subsampled (h) Activity map Laplacian subsampled (i) Activity map Laplacian oversampled. The intensity scale here is different from Figure 4-14

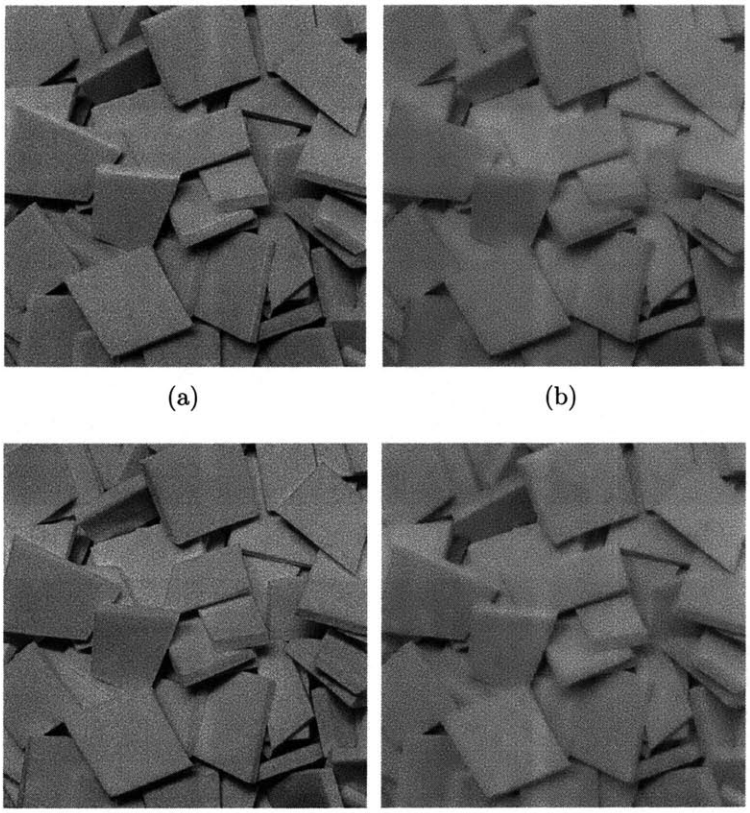


Figure 4-16: Material 2, Activity Map based Heeger-Bergen (a) B (b) R (c) R2B (d B2R)

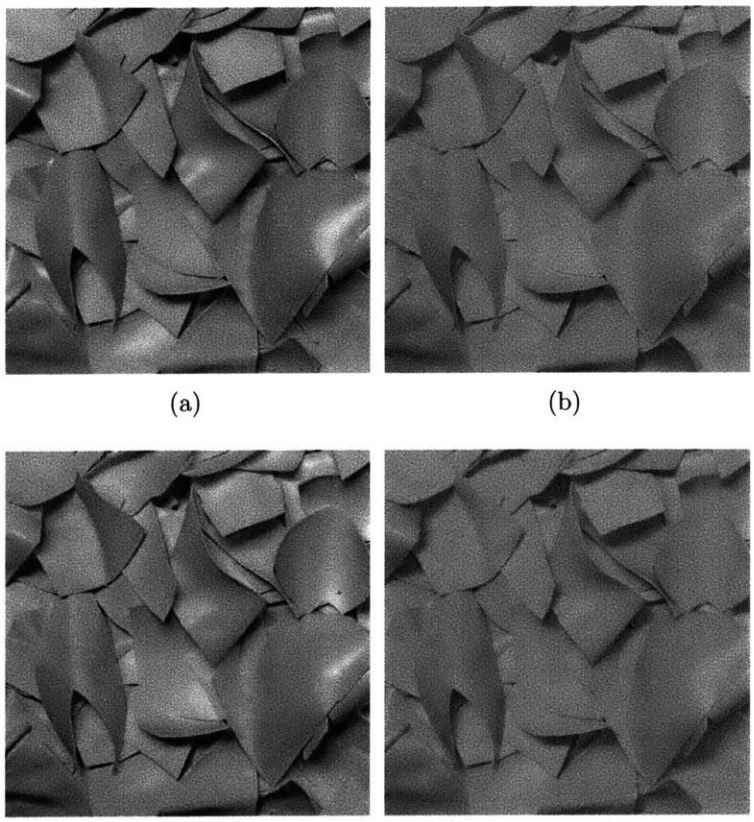


Figure 4-17: Material 3, Activity Map based Heeger-Bergen (a) B (b) R (c) R2B (d B2R)

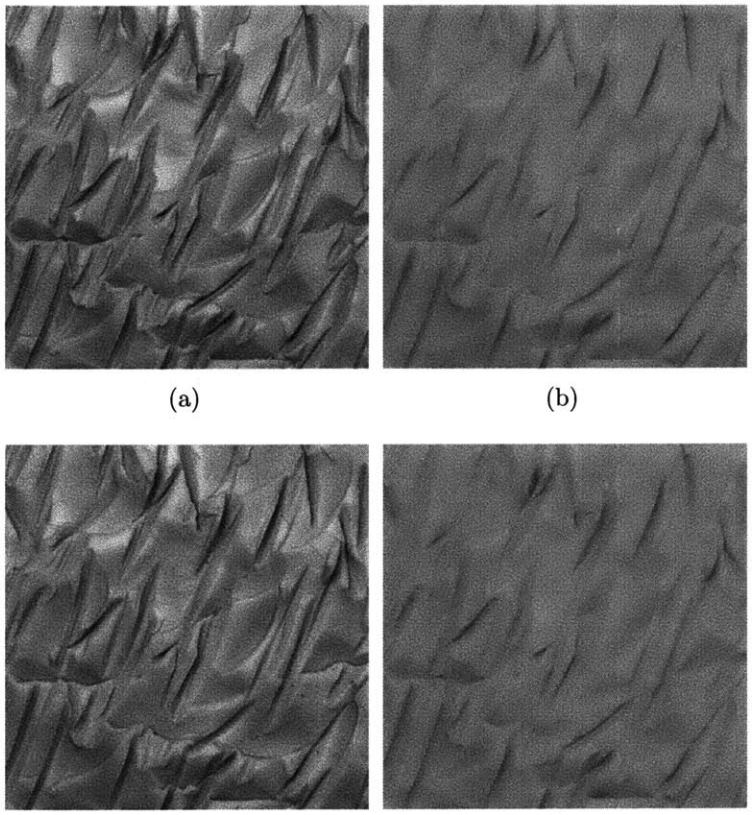
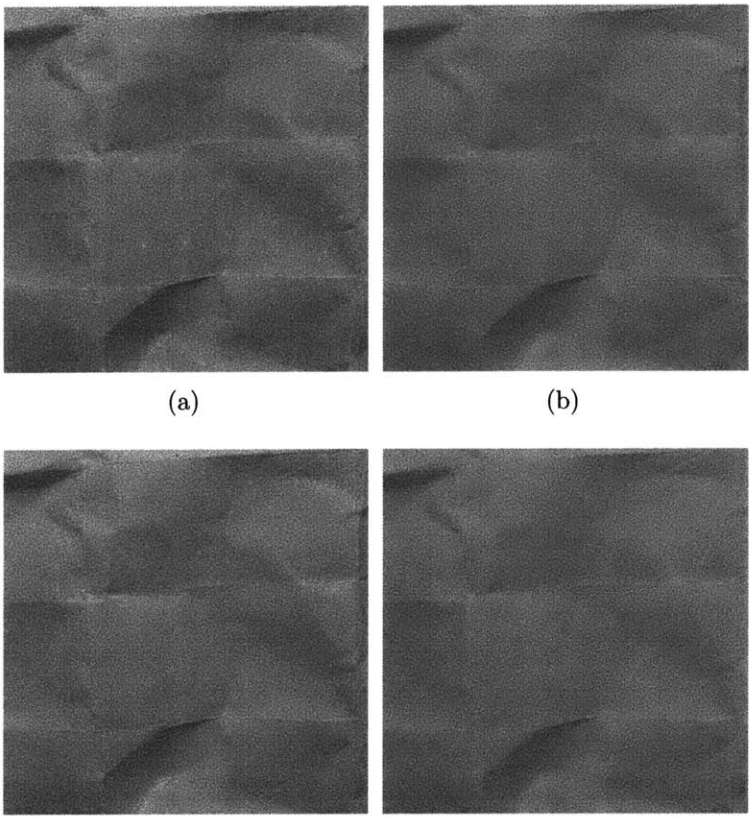


Figure 4-18: Material 5, Activity Map based Heeger-Bergen (a) B (b) R (c) R2B (d B2R)



(a) (b)

(c) (d)

Figure 4-19: Material 9, Activity Map based Heeger-Bergen (a) B (b) R (c) R2B (d B2R)

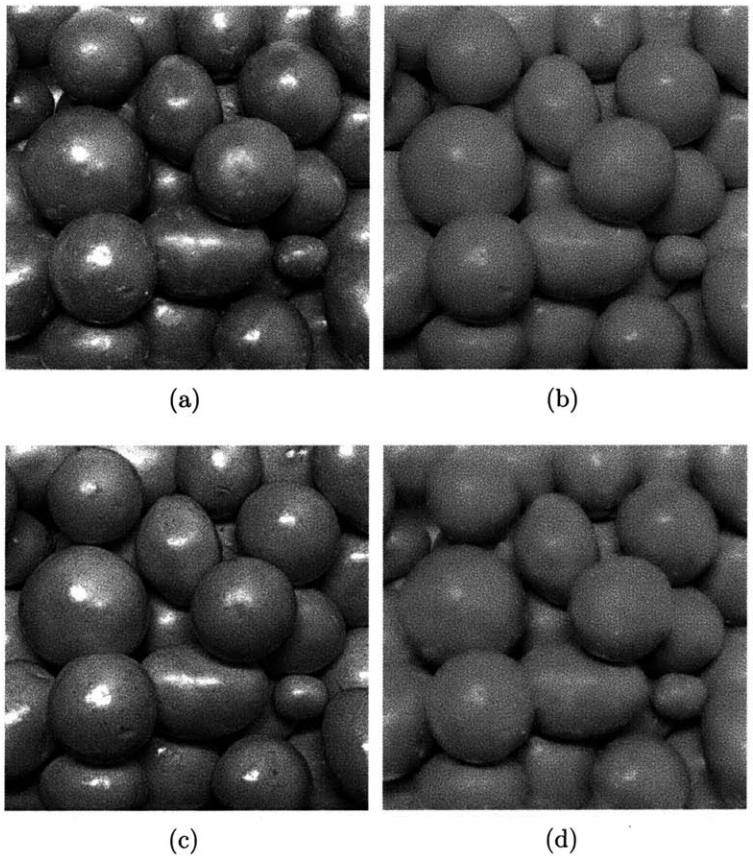


Figure 4-20: Material 14, Activity Map based Heeger-Bergen (a) B (b) R (c) R2B (d B2R)

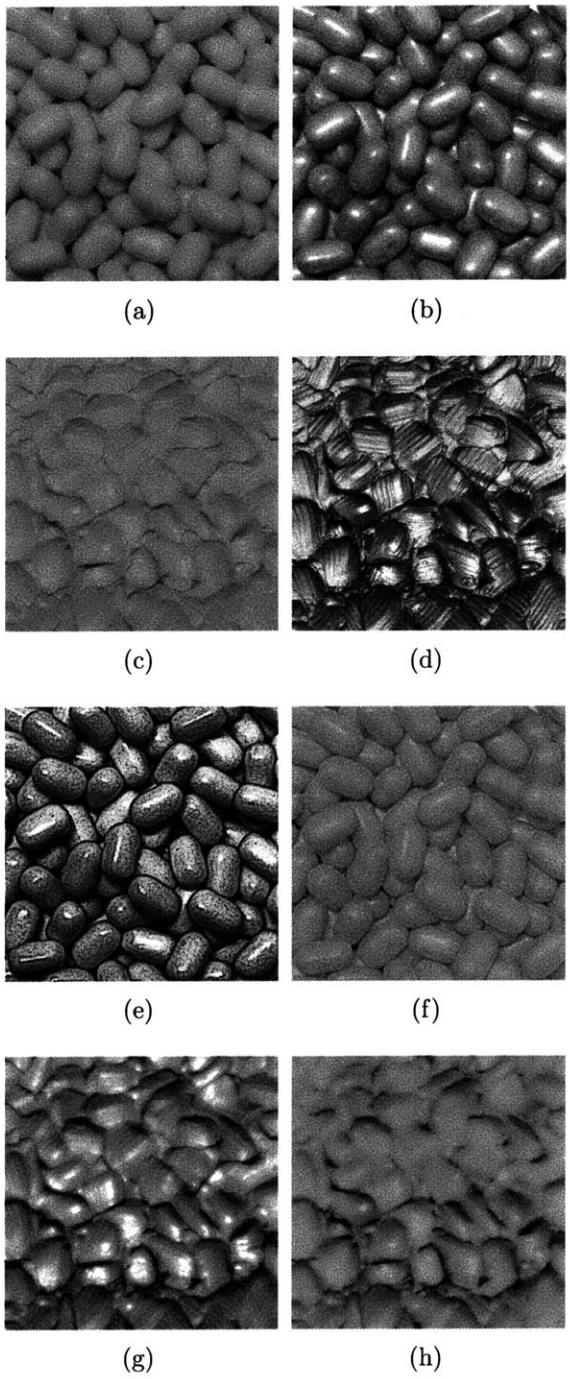


Figure 4-21: Material 1 and 10 cross, Activity map based Heeger Bergen using an oversampled Laplacian pyramid (a) Material 1 R (b) Material 1 B (c) Material 10 R (d) Material 10 B (e) Source 1 R, Target 10 B (f) Source 1 B, Target 10 R (g) Source 10 R, Target 1 B (h) Source 10 B, Target 1 R

Chapter 5

Comparison of Image Statistics with Human Performance

On examining the cases where image statistics fail to predict the reflectance (Figure 5-1), we observe that our lightness judgements about such "failures" tend to agree with the statistics than the ground truth. For example, the black foamboard in Figure 5-1a, would be categorized as a middle gray rather than black by the image statistics. Perceptually, the image does not look particularly black or dark. Figure 5-1b is the B channel of Material 3 (cut-up orange balloons). The statistics consider the reflectance of the underlying material in this image to be middle gray. If we look at the image we find that, once again, our perception parallels the image statistics.

Encouraged by this agreement between a human observer and image statistics, we want to explore how the statistics perform, compared to humans, at the task of estimating the reflectance of materials. From the anti-Gelb experiments in Chapter 1, we know that humans can distinguish materials of distinct reflectance in the absence of mean luminance information. Therefore it is meaningful to ask human observers to judge the reflectance of a material, by showing them a single mean luminance normalized image in isolation. We formulate two psychophysical experiments, the first, a 2AFC (two-alternative forced choice) classification task (Experiment I) and the second, a ratings task (Experiment II). In both experiments, we find that humans

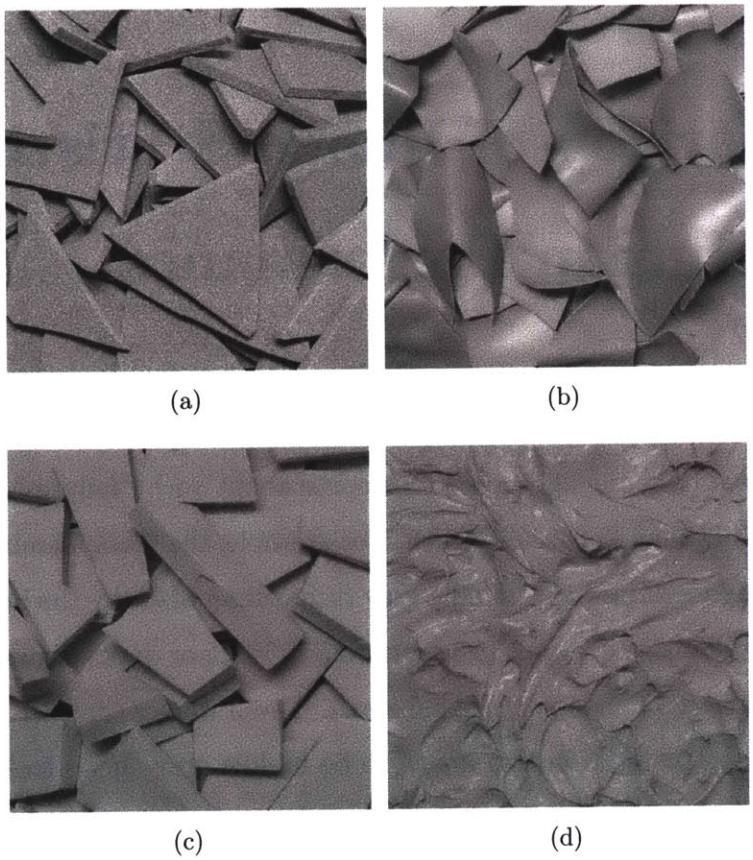


Figure 5-1: **Examples of failure of image statistics** (a) Gray material 2 (black foam board) does not look particularly black (b) Orange material 3 B channel (cut-up balloon) looks middle or dark gray, but not a convincing black (c) Orange material 4 R channel (foam board) and (d) Orange material 12 R channel (stucco) looks middle gray but not convincing white.

are not perfectly lightness constant. However the Gelb effect clearly fails for all our observers on our chosen material stimuli. Thus, human performance is somewhere between perfect constancy and no constancy. We compared the reflectance judgements of a learning algorithm that uses our image statistics, to human judgements and we find that the statistics and humans perform similarly and they succeed and fail on the same images.

5.1 Observers

Adult human observers with normal or corrected to normal vision participated in both experiments. All observers had experience participating in psychophysical experiments. Twenty-nine observers participated in the Experiment I and eighteen participated in Experiment II. Nine observers (BB, CT, KA, LS, PK, RH, SV, YL, ZC) participated in both experiments. Only two observers, LS and YL, were non-naive subjects who were aware of the purpose of the experiments. All studies were conducted in the Perceptual Science Laboratory at Massachusetts Institute of Technology.

5.2 Apparatus

All stimuli were viewed on a Dell 20.1 inch Flat Panel LCD monitor. The monitor had 1280 $\times$ 1024 resolution, 75 Hz frame rate and 70 cd/m^2 mean luminance. Observers indicated their responses by pressing appropriate keys on a keyboard. In Experiment I, they viewed the LCD in a completely dark room. In Experiment II (Figure 5-2) they could also view a box enclosing two light sources and a Munsell chart with standard reflectance chips.

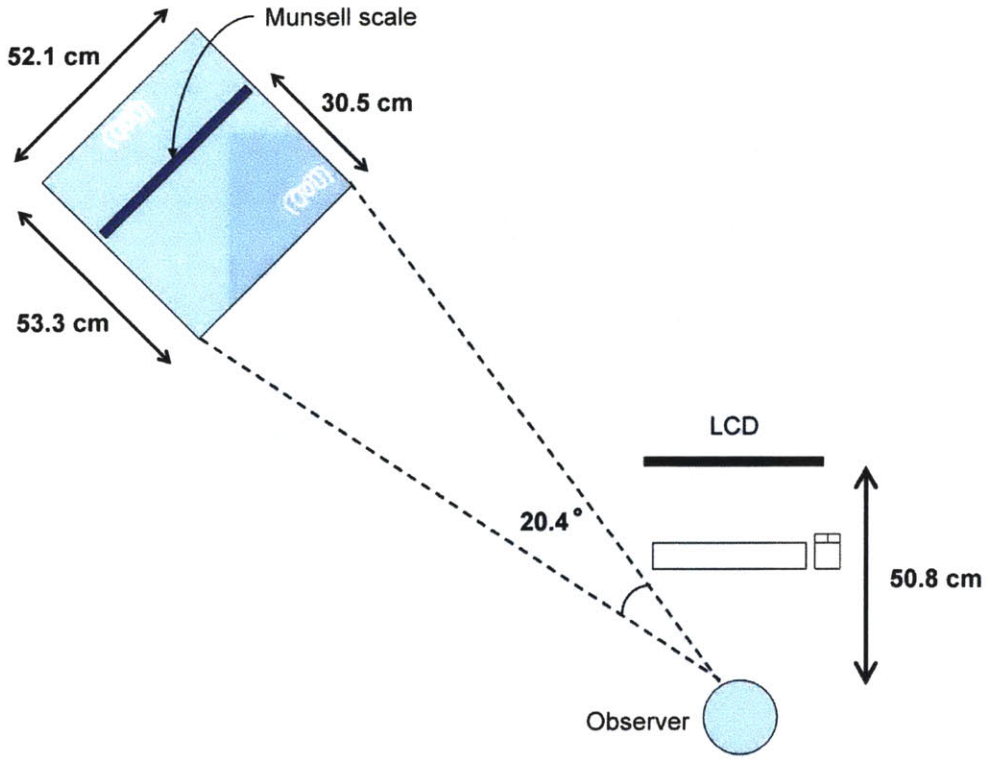
A photograph of the box with the Munsell chart is shown in Figure 5-3a. The

box was constructed from white foamboard panels and covered with dark gray craft paper on the outside. One side of the box was left open to allow the observers to see the Munsell chart. There are two small openings at the back and the side of the box (refer Figure 5-2b) for the light sources. Compact fluorescent light bulbs of color temperature 5500 K (Sunwave Full Spectrum CFL bulbs) were used to illuminate the chart. The light sources were positioned to provide approximately uniform illumination across the chart. The light sources were shaded by attaching white paper to the front of the box.

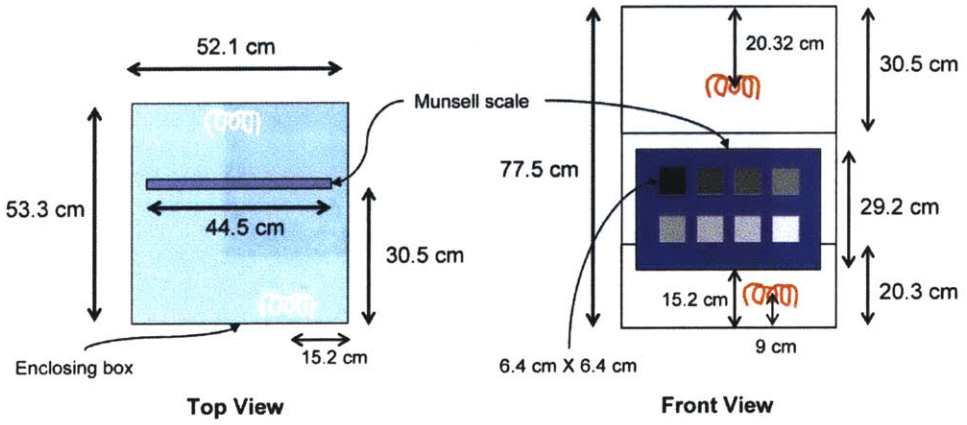
The Munsell chart (Figure 5-3b) comprised eight gray squares, numbered 1 to 8, on a random-noise-like background. The gray squares were matched to the Munsell standard reflectances N2, N3, ... N9. The squares were printed on Epson enhanced matte paper by an Epson Stylus Photo R800 printer. Each square was matched by eye to the corresponding Munsell reflectance (Gretag-MacBeth 31-step Neutral Value Scale, matte) under the Sunwave bulbs. The random-noise background was also printed on the same paper with the Epson printer. The random-noise pattern was chosen to provide a well articulated framework [23] for the gray squares.

5.3 Stimuli

The stimuli consisted of images of materials (refer Sections 3.1, 3.2 for image data acquisition and processing). Fifteen orange materials and eleven gray materials under three lighting conditions were used. The color channels of each orange material were viewed separately. Therefore, the observer always viewed a grayscale image. Figure 5-4 is an example screen shot of the stimuli the observers viewed. The images were viewed one at a time, against a constant gray background. Images were displayed so because we wanted observers to concentrate on each image independently of other images.

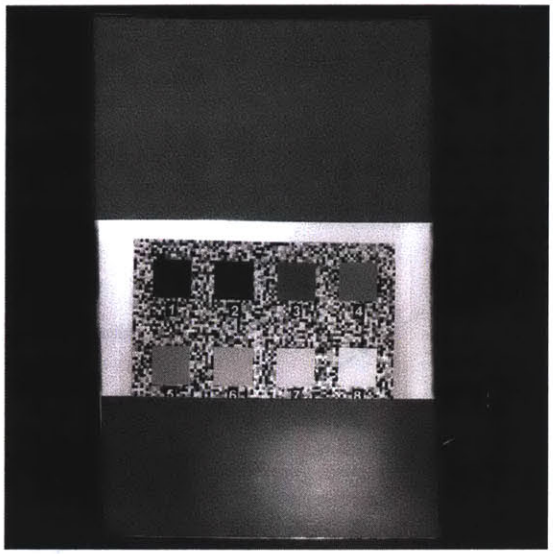


(a)

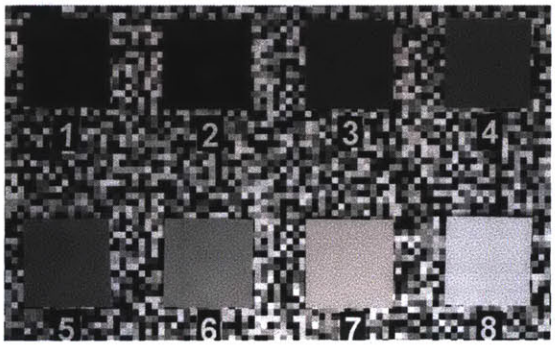


(b)

Figure 5-2: Setup for Experiment II



(a)



(b)

Figure 5-3: Munsell scale box setup

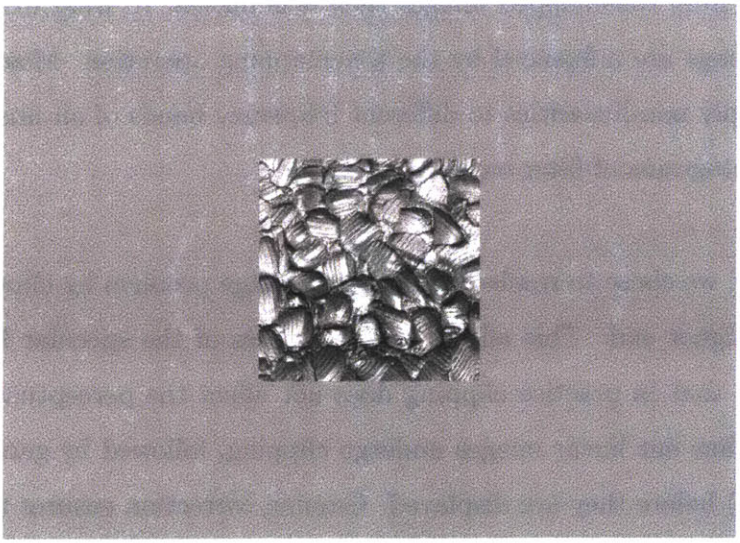


Figure 5-4: What the subjects viewed

Ideally we would like the image of a material to convey as much information as a real world 3-D viewing of the material. Scenes in the real world have a high dynamic range (ratio of highest luminance to lowest luminance). Displaying our high dynamic range (HDR) image data on a low dynamic range (LDR) device such a CRT or a LCD is a issue. Many of our materials especially the shiny ones and those with lower reflectance, have sharp specular highlights and deep shadows. In order to display images of such materials on our LDR devices, our first thought was to use HDR compression (or tonemapping) algorithms that transform HDR images to visually pleasing LDR images. Most such algorithms, however, treat each image uniquely. In other words, there is no fixed function that maps the input HDR image to the output LDR image. This is a concern for us.

We want to study the statistics that are computable directly from image intensities, because similar statistics might be employed by the visual system on the real world luminances. By using an input-dependent HDR compression scheme we lose access to the original intensities in the image. Moreover, if humans view tonemapped images and image statistics are calculated directly from the image intensities then a comparison of human performance with that of statistics is meaningless. If statistics

are not perfectly lightness constant. However the Gelb effect clearly fails for all our observers on our chosen material stimuli. Thus, human performance is somewhere between perfect constancy and no constancy. We compared the reflectance judgements of a learning algorithm that uses our image statistics, to human judgements and we find that the statistics and humans perform similarly and they succeed and fail on the same images.

5.1 Observers

Adult human observers with normal or corrected to normal vision participated in both experiments. All observers had experience participating in psychophysical experiments. Twenty-nine observers participated in the Experiment I and eighteen participated in Experiment II. Nine observers (BB, CT, KA, LS, PK, RH, SV, YL, ZC) participated in both experiments. Only two observers, LS and YL, were non-naive subjects who were aware of the purpose of the experiments. All studies were conducted in the Perceptual Science Laboratory at Massachusetts Institute of Technology.

5.2 Apparatus

All stimuli were viewed on a Dell 20.1 inch Flat Panel LCD monitor. The monitor had 1280 $\times$ 1024 resolution, 75 Hz frame rate and 70 cd/m^2 mean luminance. Observers indicated their responses by pressing appropriate keys on a keyboard. In Experiment I, they viewed the LCD in a completely dark room. In Experiment II (Figure 5-2) they could also view a box enclosing two light sources and a Munsell chart with standard reflectance chips.

A photograph of the box with the Munsell chart is shown in Figure 5-3a. The

humans are assumed to be trained and indicate the reflectance by comparing to a standard scale.

In both experiments, observers viewed grayscale images of materials against a middle gray background on an LCD display. All images were normalized to have the same mean luminance and were displayed one at a time. Observers could take as long as they wanted to judge the reflectance of an image. The inter stimulus interval was fixed at 0.5 seconds. For the first experiment, observers sat in a dark room, where the only object they could view was the LCD screen. Observers indicated the category of the material by pressing appropriate keys on the keyboard. In the second experiment, observers could look back and forth between the LCD and the Munsell chart with standard reflectance squares. Observers were asked to provide a number between 1 and 8 to indicate the square on the chart they believed to be closest in reflectance to the material on screen. Fractional ratings like 4.5 were permitted. This allowed observers to express their confidence about the rating.

For both experiments, observers were instructed to judge the lightness and not the overall brightness of an image. Observers were presented a practice trial before they proceeded to the main experiment.

5.5 Experiment I

5.5.1 Motivation

We formulate the first experiment as a two alternative forced choice (2AFC) classification task. Images are divided into two categories - light and dark. We assume that observers are not familiar with our stimuli hence they have to undergo training. In the training phase observers learn the relationship between the category labels and the training images. Once they move to the test phase, they can classify test images

with ease. This experiment was designed to allow a fair comparison of humans with a learning algorithm that uses images statistics as features. Both the learning algorithm and observers treat each image independently of others, and undergo a training and a test phase.

5.5.2 Procedure

Observers viewed images of mean normalized materials on an LCD display in a dark room. The images were divided in two categories - A, where the ground truth for reflectance is > 0.5 (light gray to white) and B, reflectance < 0.2 (dark gray to black). The images were displayed against a gray background of intensity $= 0.33$ where 1 is the maximum intensity of the display (Figure 5-4). Observers started with a training phase, where they viewed each image one at a time, for as long as they wanted till they labelled the image as Category A or B. For each training image, observers were provided immediate feedback on their performance. On completing the training phase they moved to the test phase. The test phase is identical to the training phase except that there is no feedback.

In the training phase, observers viewed 10 materials (6 orange) under three illumination conditions and in the test phase they viewed 20 materials (13 orange) under three illumination conditions. The training and test set materials were disjoint. The experiment was run with 29 subjects. Observers were divided into three groups (Group 1, 10 observers, Group 2, 9 observers and Group 3, 10 observers). Each group viewed different training-test set pairs. For each observer and each material, 2 repetitions were run per illumination condition. For orange materials, there were 2 repetitions per condition for each channel (R or B). The order of the images was randomized.

5.5.3 Results

The results of experiment I are demonstrated in a bar plot in Figure 5-5 in blue. Data is pooled across all observers and all lighting conditions. The percentage of images correctly classified by observers is plotted for each group both for the training phase and the test phase. Error bars indicate the 95% confidence intervals of the mean. Percentage correctly classified is calculated as follows - each image (a combination of material and light and if applicable channel) is counted if it was correctly classified both times it was viewed by an observer.

The individual performance of each observer is recorded in Table 5.1. Both from the table and the bar plot we observe that most subjects perform with an accuracy of $75\% - 85\%$ both on the training set and the test set. This performance is way above chance (50%) and indicates that the classification task is meaningful and that observers can do it easily. The performance of all three groups is similar, this may be because the materials were divided in groups carefully. Some materials are harder to judge than others but about the same number of *hard* materials were included in each group.

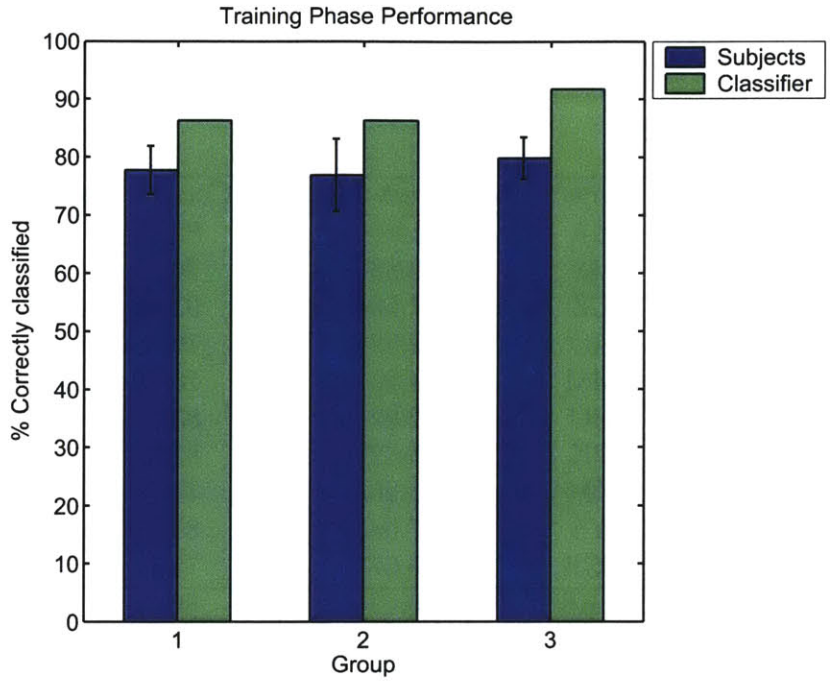
To compare this performance with image statistics, we choose a simple linear regression classifier and train it with three image statistics - $(90^{th} - 10^{th})$ percentile of the gaussian center surround filter output at the finest level, skew of the intensity values, 99^{th} percentile of the Sobel filter output at the finest level. The classifier is trained on the same training images as humans for each group, and then tested on the same test set. The classifier performance is illustrated in the bar plot (Figure 5-5) in green. The performance of the classifier is comparable to that of the observers, both on the training and the test sets. This is an encouraging and somewhat unexpected result given the simplicity of the learning technique and the fact that only three image statistics were chosen.

A problem with the design of this experiment is that in order to classify images, observers may not be judging reflectance at all, but some other attribute that distinguishes the stimuli of the two categories. Moreover we used both channels R and B for each orange material and each channel belongs to a different category. Once the observers figure this out, they can classify the B channel correctly once they have seen the R and vice-versa. The reason for including both the R and B channels of the same material in each group was to ensure the type of material does not affect the observer's judgments (e.g. stuccos are category A and TicTacs category B).

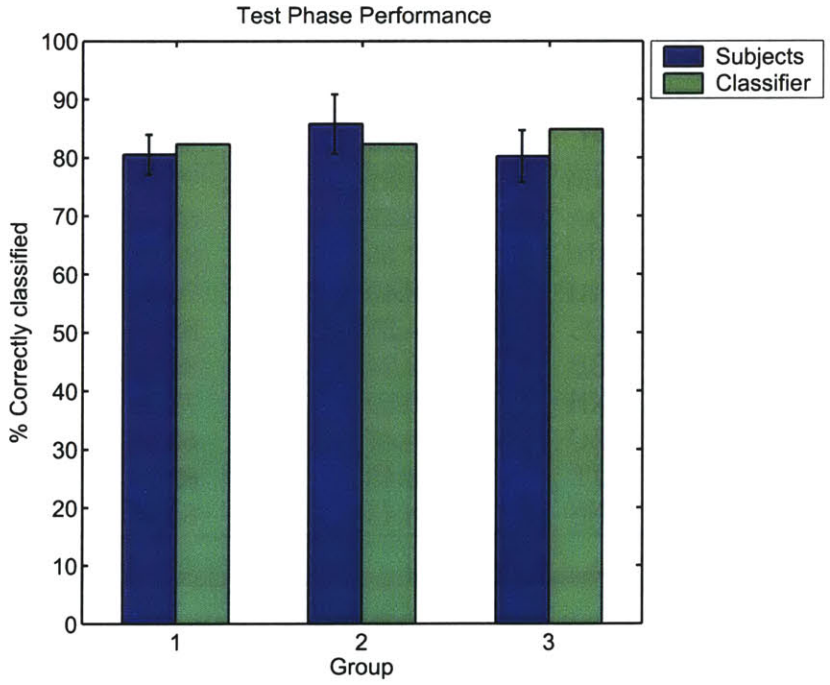
Given that, the observers and the classifier perform equally well at the classification task, it is interesting to ask if they make the same mistakes? Are the 20-odd percent images that both get wrong the same? Figure 5-6 shows examples of the errors made by subjects and the classifiers. There is clearly an overlap in the errors (69% of all the images that both get wrong). However the classifier does make mistakes that observers do not (Figure 5-6e,f) and these mistakes are highly sensitive to the features that are used by the classifier. Finally, we find that the lighting condition does not affect the classification performance. If observers label a material correctly, they do so for all the three lights.

Group	Subject	% Correct Train	% Correct Test
1	KA	81.25	82.83
	MF	70.83	74.75
	CZ	87.50	91.92
	AC	72.92	79.80
	MM	81.25	78.79
	BC	72.92	83.84
	PK	85.42	81.82
	XM	81.25	77.78
	YS	77.08	81.82
	JCH	66.67	71.72
2	BHS	85.42	88.89
	JC	83.33	81.82
	RH	83.33	90.91
	ACH	85.42	79.80
	SV	66.67	74.75
	JA	77.08	87.88
	MG	64.58	76.77
	SS	83.33	95.96
	SL	62.50	94.95
3	VM	84.31	77.08
	DA	86.27	85.42
	MH	72.55	83.33
	DAP	78.43	76.04
	DS	84.31	88.54
	BB	88.24	84.38
	XH	70.59	77.08
	ZC	76.47	63.54
	JT	78.43	82.29
	NS	78.43	84.38

Table 5.1: **Experiment I results** Percentage correctly classified for each subject for training and test phases

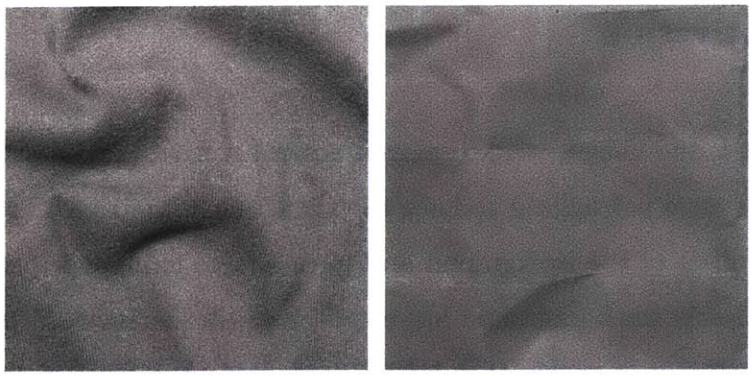


(a)

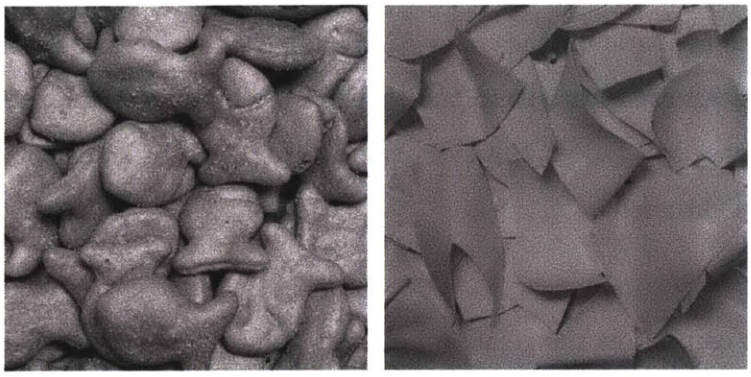


(b)

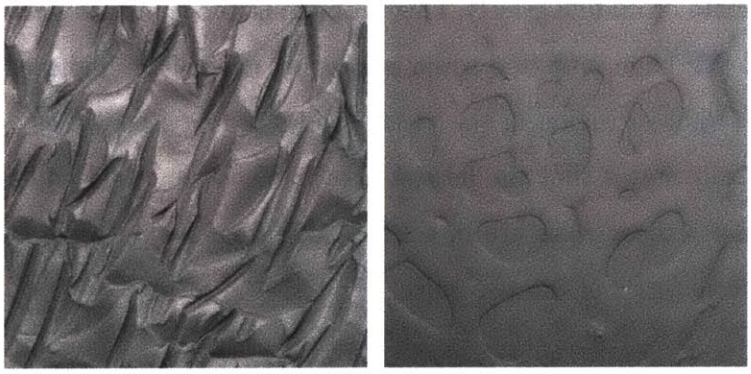
Figure 5-5: Performance of observers and a linear regression classifier based on image statistics on the classification task. The error bars are the 95% confidence intervals



(a) B classified A (b) A classified as B



(c) B classified as A (d) A classified as B



(e) B classified as A (f) B classified as A

Figure 5-6: Errors by observers and classifier on Experiment I (a)-(b) Errors only made by observers (c)-(d) Errors made by both observers and classifier (e)-(f) Errors only made by the classifier

5.6 Experiment IIA

5.6.1 Motivation

In experiment I, the images were displayed against a fixed middle gray background and all images were normalized to have the same mean luminance. We want to investigate the effect of the background and mean image luminance on the perceived reflectance. Clearly the background has no effect on the statistics of the image being displayed, however it is known that humans use contextual cues in lightness judgements. Wallach [53] formulated a *ratio rule.* According to this rule, the visual system measures ratios of the luminance of spatially adjacent surfaces to estimate lightness. Because the ratios are invariant to the absolute illumination level, the visual system achieves lightness constancy. However, the ratio rule is not sufficient to explain the perception of lightness.

The ratios can tell us the relative lightness values of the surfaces however we do not know the absolute lightness values (e.g. is the surface with the higher luminance perceived as white or middle gray?) This problem is known as the *anchoring problem.* Gilchrist et al [23] have proposed some *anchoring rules.* The highest luminance rule says that the brightest region in the visual field will be perceived as white. Another rule, the area rule states that the largest area in the visual field will be perceived as white. Sometimes these rules are in conflict and what we perceive is a compromise between the predictions of the two rules.

To explore contextual effects further, we separate Experiment II into IIA, IIB and IIC. In Experiment IIA, we study the effect of the luminance of the gray background, mean luminance of the image and their ratio on the perceived reflectance of the material.

5.6.2 Procedure

Observers viewed one particular orange material (all channels R,G, and B) under three lights for different conditions of background luminance, mean image luminance and ratio of image to background luminance. The experiment was divided into three sections. In the first section, the luminance of the gray background was varied through $[0.11, 0.17, 0.33, 0.67, 1]$ where 1 was the maximum luminance of the display. The mean luminance of the image was held fixed at 0.33. For each channel (R,G or B) 2 repetitions were run per lighting condition for each level of background luminance. In the second section, the background luminance was held fixed at 0.33 while the mean luminance of the image was varied through $[0.1, 0.2, 0.3, 0.35, 0.4]$. As in the first section, for each channel (R,G or B) 2 repetitions were run per lighting condition for each level of image luminance. In the third section, the luminance ratio of the background and the image was held fixed at 1 and the mean luminance of the image/background was varied through $[0.1, 0.2, 0.3, 0.35, 0.4]$. As before, for each channel (R,G or B) 2 repetitions were run per lighting condition for each level of screen luminance. The order of images was randomized within each section.

Eight observers participated in this experiment. Observers were divided into two groups, four in each. Observers in Group 1 viewed images of orange Material 5 and those in Group 2 viewed orange Material 10 (Figure 5-7). This experiment lasted about 40 minutes.

5.6.3 Results

Section 1

In this section, observers viewed images of Material 5 (orange modelling clay) or 10 (orange stucco) in three lights and five different conditions of the background luminance. The mean image luminance was held fixed at 0.33. If observers are perfectly lightness constant, their responses would be the same for all conditions of the back-

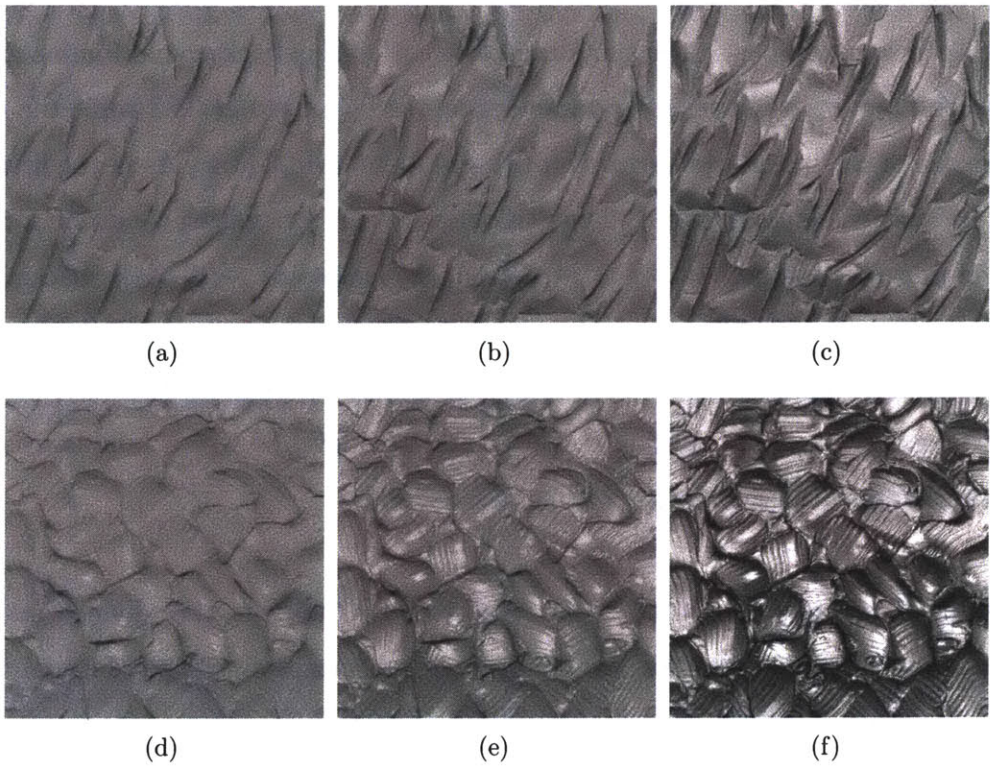


Figure 5-7: Stimuli for Experiment IIA (a)-(c) R,G, B channels of Material 5, (d)-(f) R,G, B channels of Material 10

ground luminance. If observers display no constancy and follow the ratio rule [53] in conjunction with an anchoring rule [23] then their responses should be a function of the background luminance. Figure 5-8 plots the perceived reflectance versus the background luminance on a log-log scale for each observer for R channel of Material 10. We find that for all observers the perceived reflectance lies somewhere between perfect constancy and no constancy. As the luminance of the background increases, the perceived reflectance decreases as a non-linear function of the background luminance. The lighting condition has a slight effect on the responses but not too much. The same observations hold for the G channel of Material 10 (Figure A-2) and all channels of Material 5 (Figures A-4 through A-6). However for the B channel of Material 10 (Figure 5-9) for all observers except KA, the background luminance does not affect the perceived reflectance. This does not mean that observers are perfectly lightness constant as all three of them overestimate the true reflectance of the material. Thus, the results in this section lead us to conclude that the luminance of the background influences the lightness judgements.

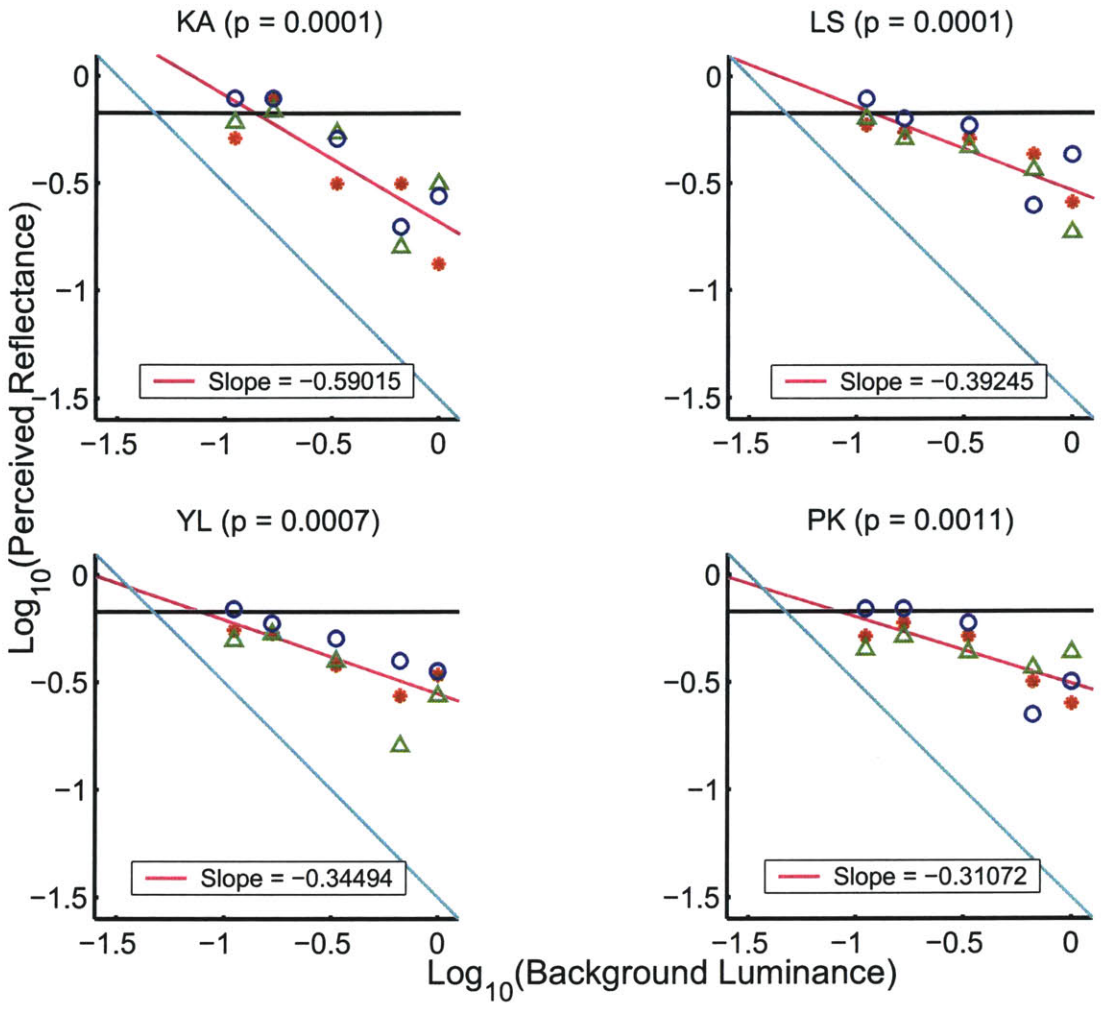


Figure 5-8: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Background luminance) (Material 10, R channel)**. Luminance of material image is held fixed at 0.33 while the luminance of the background changes. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log background luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio rule, the responses would lie along a line parallel to the cyan line with slope = -1 . The magenta line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers the slope of the fit is significantly different from 0 and -1.

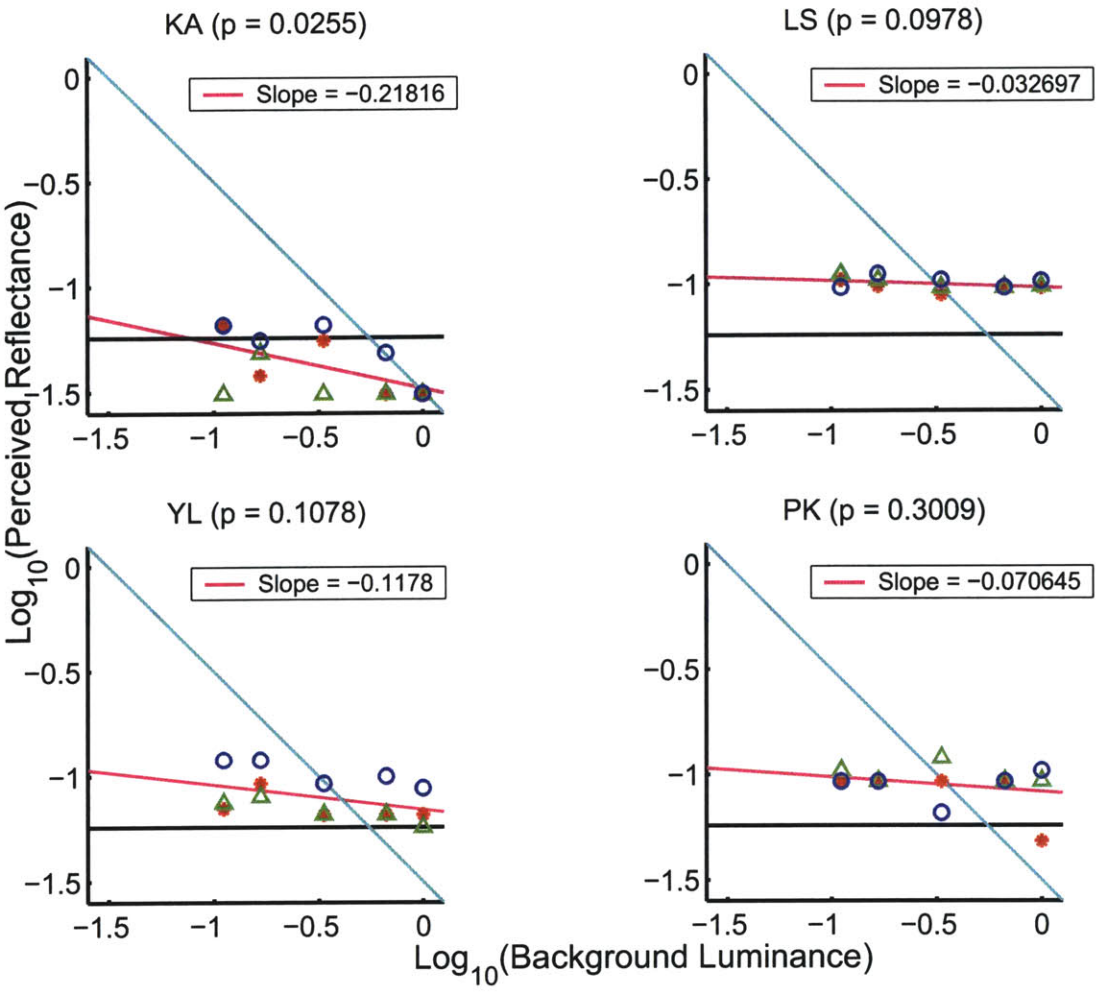


Figure 5-9: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Background luminance) (Material 10, B channel)**. Luminance of material image is held fixed at 0.33 while the luminance of the background changes. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log background luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio rule, the responses would lie along a line parallel to the cyan line with slope $= -1$. The magenta line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers, except KA, the slope of the fit is not significantly different from 0.

Section 2

In this section, the background luminance was held fixed at 0.33 while the mean luminance of the image was varied. As before, if observers are perfectly lightness constant their responses will be the same for all five conditions of the mean luminance. If they exhibit zero constancy and follow the ratio rule their responses will be a function of the mean image luminance. Figures 5-10 and A-8 through A-12 plot the perceived reflectance versus mean image luminance on a log-log scale. For nearly all observers and all channels of Materials 5 and 10, the responses lie between perfect constancy and zero constancy. As the mean luminance of the image increases, the perceived reflectance increases as a non-linear function. For some cases, the responses of the observers align with the no constancy line and in general the slopes of the fit to observers' responses are closer to 1 (no constancy) in this section, as compared to the previous section. This suggests that the mean image luminance is important for lightness judgements and perhaps more so than the luminance of the background.

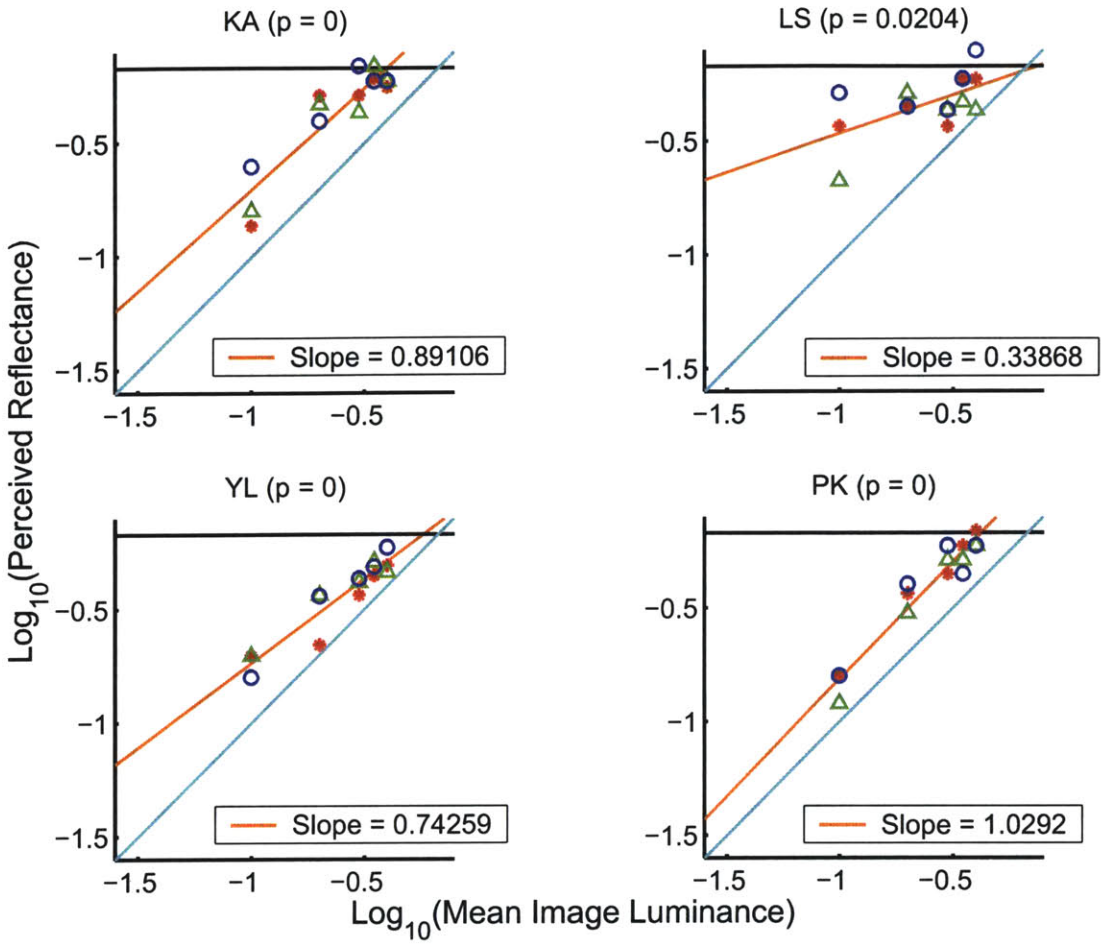


Figure 5-10: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Mean Image luminance) (Material 10, R channel).** Luminance of background is held fixed at 0.33 while the mean luminance of the material image changes. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log mean image luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio rule, the responses would lie along a line parallel to the cyan line with slope = 1 . The orange line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers, except PK, the slope of the fit is significantly different from 0 and 1.

Section 3

In this section, the ratio of the mean image luminance to background luminance was held fixed at 1 while the mean image/background luminance was varied. As before, if observers are perfectly lightness constant their responses will be the same for all five conditions of the screen luminance. If they exhibit zero constancy and follow the ratio rule (and anchor to something outside the LCD framework) their responses will be a function of the screen luminance. Figures 5-11 and A-14 through A-18 plot the perceived reflectance versus screen luminance on a log-log scale. For all observers and all channels of Materials 5 and 10, the responses lie between perfect constancy and zero constancy. As the screen luminance increases, the perceived reflectance increases as a non-linear function. These results suggest both the mean image luminance and the background luminance affect the perceived reflectance however their interaction is more complex than a simple ratio. If only the ratio mattered, then the responses would not be a function of the screen luminance.

To analyze the effect of the luminance ratio on perceived reflectance, data was pooled across sections 1,2 and 3. Figures 5-12 and A-20 through A-24 plot perceived reflectance versus luminance ratio. For all observers and both materials, it is the case that the luminance ratio has a significant effect on the perceived luminance. This supports the conclusion that both the mean luminance and the background luminance affect the perception of reflectance, however it is not clear how these two factors interact.

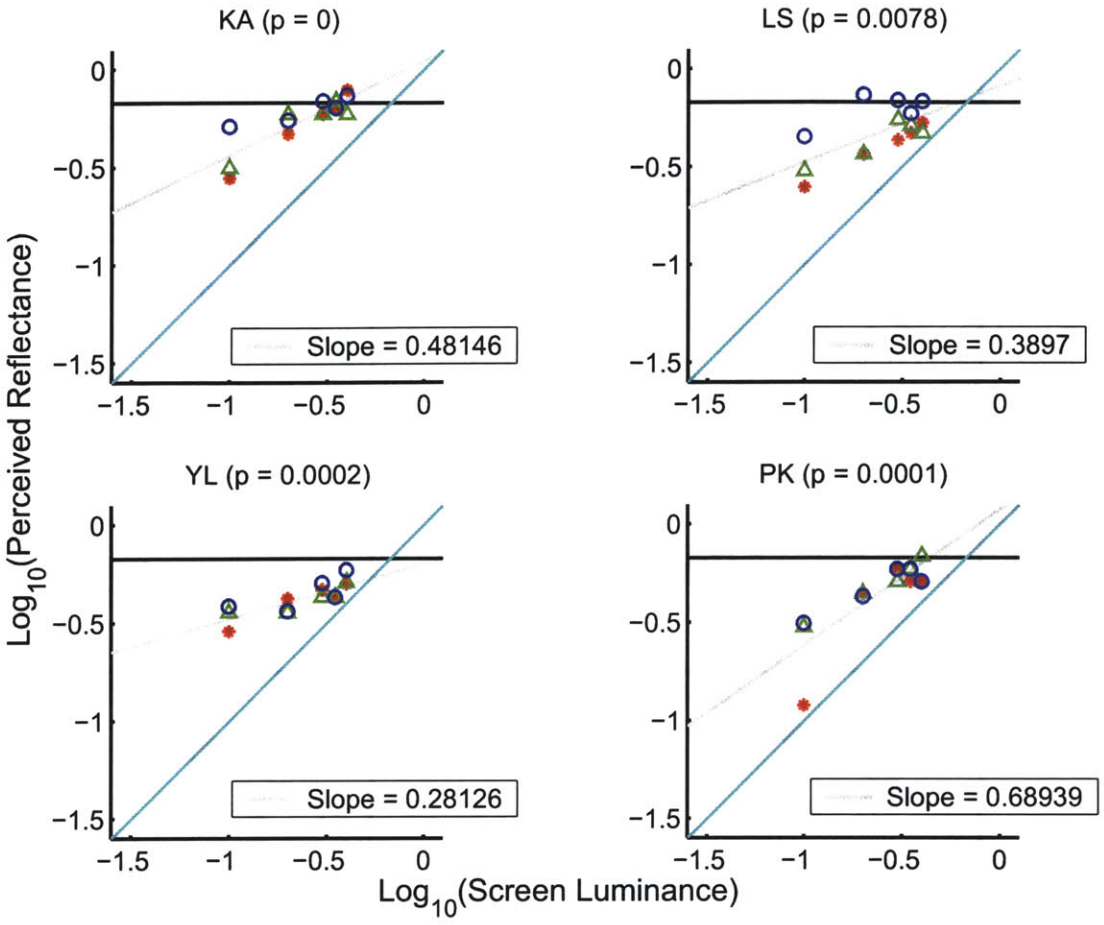


Figure 5-11: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Screen luminance) (Material 10, R channel)**. The ratio of the luminance of background to that of the image is held fixed at 1 while the mean luminances of the material image and background change. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log screen luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio rule, the responses would lie along a line parallel to the cyan line with slope = 1 . The gray line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers the slope of the fit is significantly different from 0 and 1.

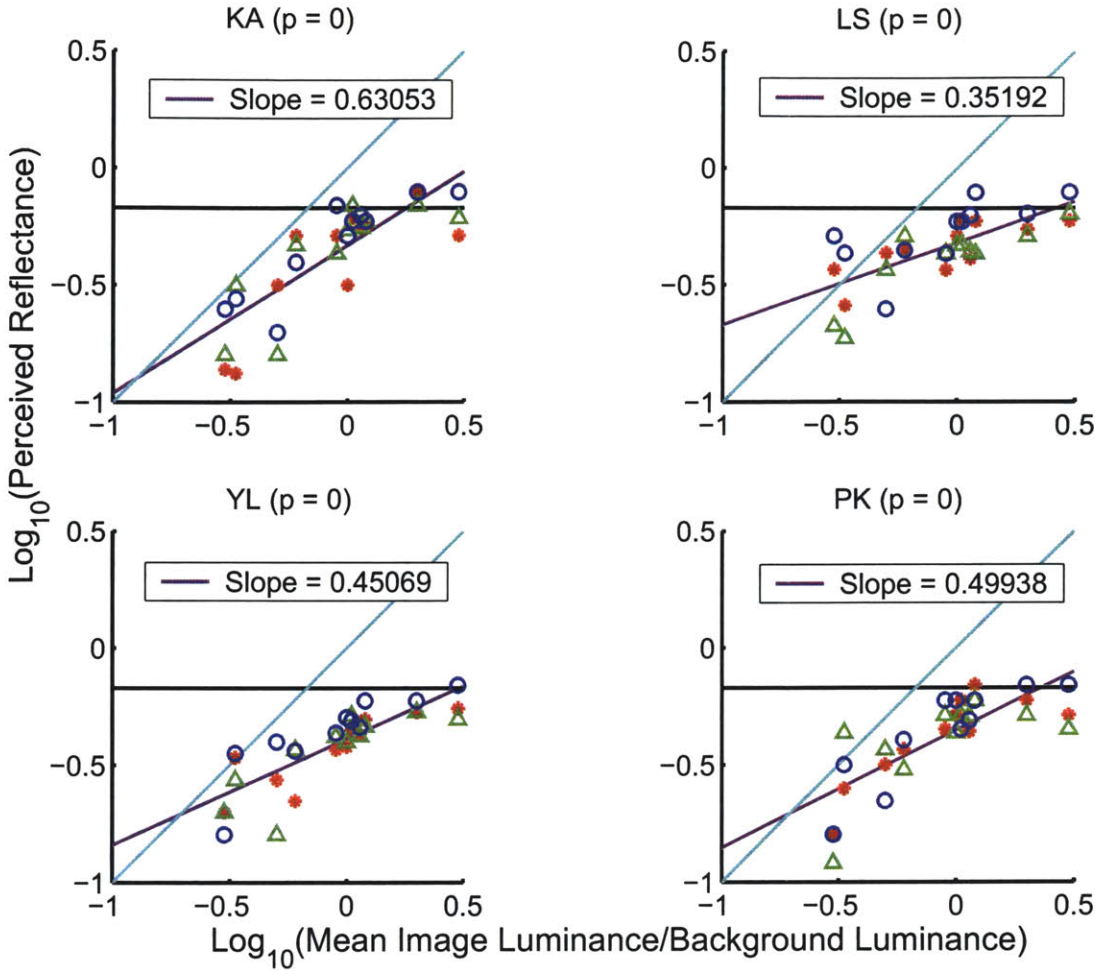


Figure 5-12: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Luminance Ratio) (Material 10, R channel).** The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log of the ratio of image luminance to background luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio rule, the responses would lie along a line parallel to the cyan line with slope = 1 . The purple line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers, the slope of the fit is significantly different from 0 and 1.

The results of Section 1,2 and 3 demonstrate that the mean image luminance and the background luminance affect the perceived reflectance of a material. It makes sense to ask if these findings negate the anti-Gelb observations we made earlier. The answer is no. While the mean image luminance and the background affect the perceived reflectance of a material in non-linear way, they preserve any ordering of reflectance. Consider Figure 5-13. The perceived reflectance is plotted against changing variable for all observers for Material 10. Note how the perceived reflectance for the R channel is always higher than that for the G channel which in turn is higher than the B channel. This is true for all conditions and all observers. The error bars indicate the 95% confidence intervals. Figure 5-14 shows a similar plot for Material 5. In this case as well, the ratings are in the order $R > G > B$ however they are not as differentiated as Material 10. These plots demonstrate that for identical conditions, the three channels of orange materials are consistently rated in the correct order of reflectance. Does this extend to the case where the different channels of different materials are viewed under identical conditions? This question is answered in Experiment IIC.

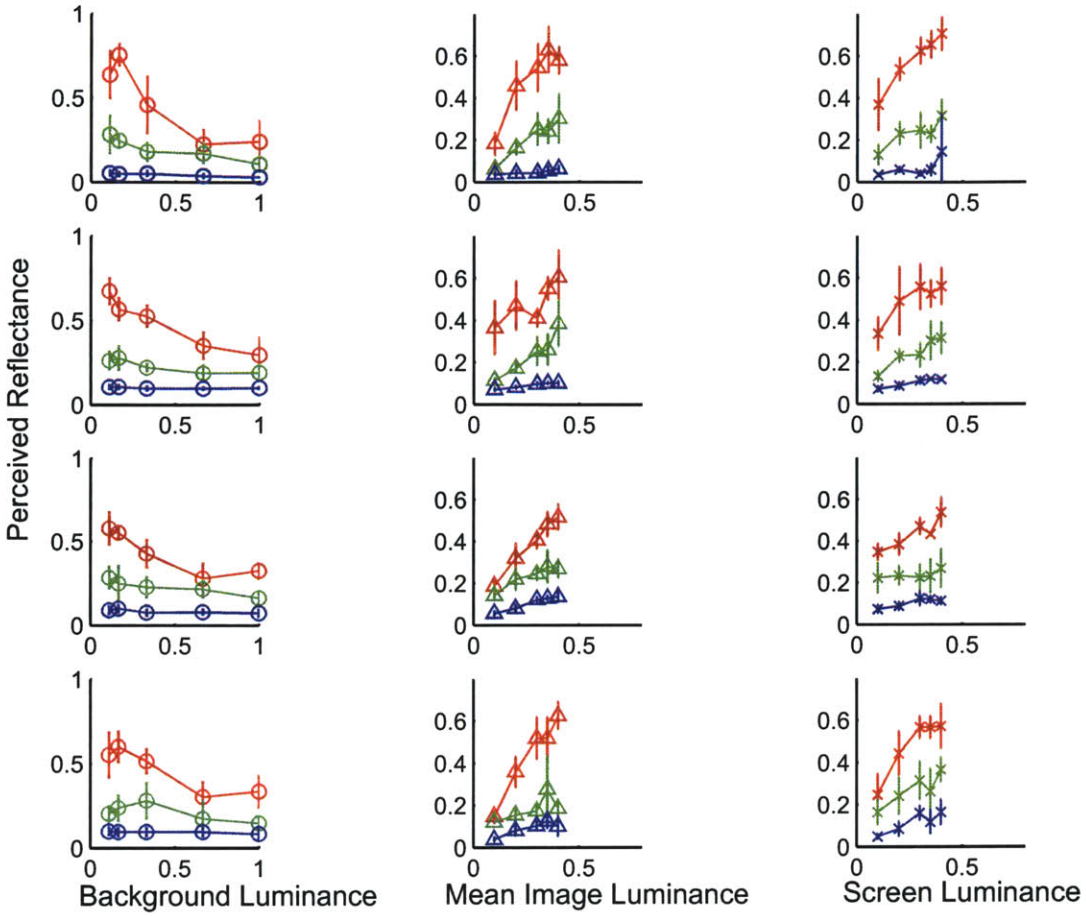


Figure 5-13: **Material 10** Perceived reflectance is plotted against background luminance (keeping mean image luminance constant) in column 1, against mean luminance (keeping background luminance constant) in column 2 and against the screen luminance (keeping the ratio of image to background constant) in column 3. Each row corresponds to each of the 4 observers (KA, LS, PK and YL) who participated in these tasks. In each plot, the R channel is denoted by red, the G channel by green and B channel by blue. For each channel, the data is pooled across the 3 lighting conditions. The errorbars are the 95% confidence intervals. While the observers do not display perfect lightness constancy, they can nevertheless differentiate between the R, G and B channels for identical experimental conditions. They consistently rate the channels in the order $R > G > B$ or as in the plot $red > green > blue$.

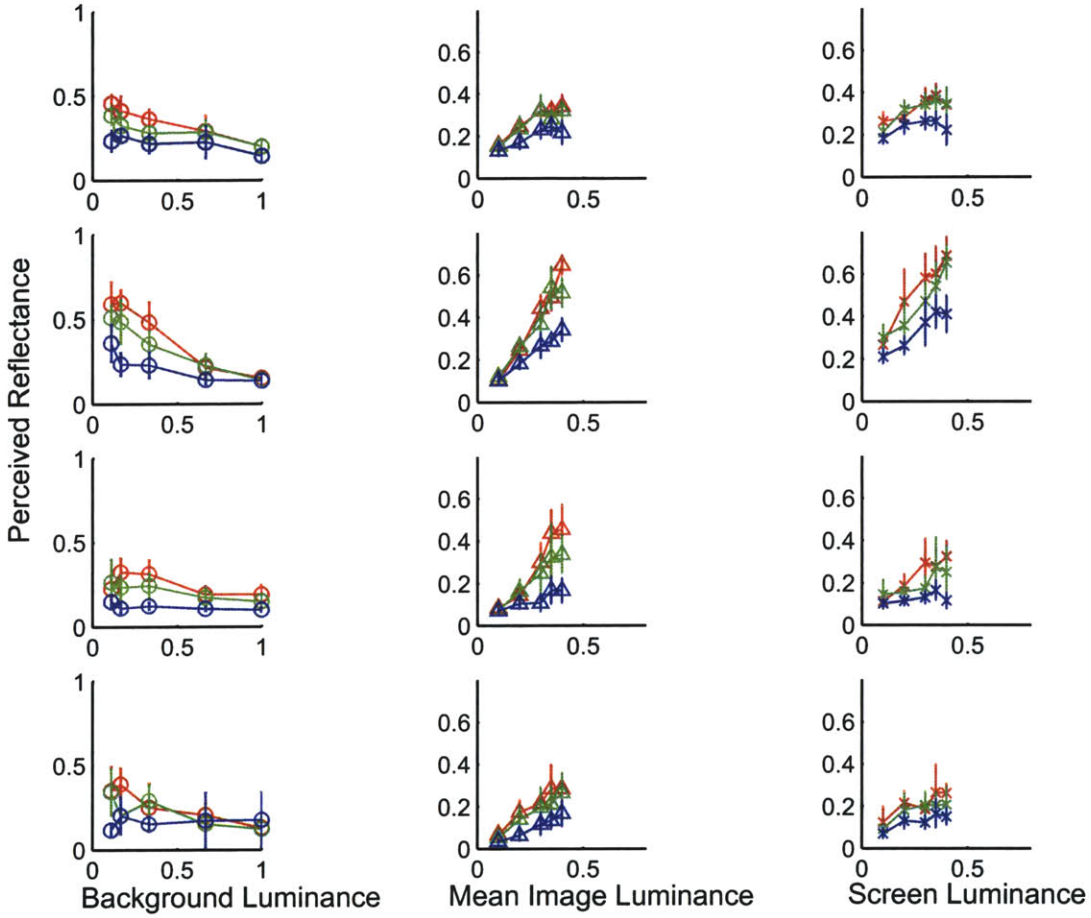


Figure 5-14: **Material 5** Similar plot to Figure 5-13, except that observers viewed Material 5 instead of Material 10.

5.7 Experiment IIB

5.7.1 Motivation

In Experiment IIA we found that for black materials viewed in isolation (Figure 5-9) humans display some constancy; this contradicts the ratio rule. For white materials (Figure 5-8) our observers display more constancy than what is predicted by the ratio rule. We want to relate these results to previous work in lightness perception. In our experiments we use complex visual stimuli like images of real world surfaces. Traditionally most studies in lightness perception (including those of Wallach's [53] and Gilchrist et al's [23]) have used flat Lambertian surface patches. Perhaps, the results of Experiment IIA do not conform to the ratio rule because we used complex visual stimuli. In order to test this hypothesis, we decided to repeat Experiment IIA but with simple visual stimuli similar to those used in previous studies.

5.7.2 Procedure

The setup for Experiment IIB was identical to that of Experiment IIA except that observers viewed blank gray patches on the LCD instead of photographs of materials. This experiment serves as a control for Experiment IIA. In the absence of any textural cues, we expected observers to follow the ratio rule [53] in conjunction with an anchoring rule [23]. Observers were instructed to match the gray level of the square in the center to one of the eight standard patches.

Experiment IIB had four sections. In the first section, the luminance of the background was varied through $[0.11, 0.17, 0.33, 0.67, 1]$ while the image in the center was a gray square of luminance 0.33. Two repetitions were run for each background luminance level. In the second section, the luminance of the gray square in the center was varied through $[0.11, 0.17, 0.33, 0.67, 1]$ while the luminance of the gray background was held fixed at 0.33. Two repetitions were run for each image lumi-

nance level. In the third section, the ratio of the gray square in the center and the gray background was held fixed at 0.5 and the square luminance was varied through $[0.11, 0.17, 0.33, 0.67, 1]$. Two repetitions were run for each luminance level. In the fourth section, the ratio of the square luminance to background luminance was held fixed at 2 and the background luminance was varied through $[0.11, 0.17, 0.33, 0.67, 1]$. Two repetitions were run for each luminance level. Ten observers participated in this experiment. This experiment lasted 7 minutes on average.

5.7.3 Results

Blank gray patches offer no textural cues; therefore we expect observers to display zero constancy and follow the ratio rule in conjunction with an anchoring rule. Figures 5-15 and 5-16 plot the pooled responses of observers against the luminance ratio of the patch and the background (magenta). For comparison, the data from Experiment IIA are also plotted (R channel is red, G channel is green, B channel is blue). If the observers follow the ratio rule while judging lightness, their responses would lie along the black line with slope = 1. In Figure 5-15 the slope of the magenta line is statistically different from 1. For low values of the background luminance (high values of patch to background luminance ratio) the perceived reflectance levels off. In Figure 5-16 the slope of the magenta line is close to 1 though it is still significantly different from 1.

In Figure 5-17 the luminance ratio of the patch and the background is fixed. If observers follow the ratio rule, then the perceived reflectance should not change as the absolute luminance of the patch increases or decreases i.e. the observations should lie along the horizontal black line. The magenta lines in Figure 5-17 have a slope statistically different from 0. From Figures 5-15 to 5-17, we make two observations - observers' ratings for the simple stimuli like gray patches are not qualitatively different from those for complex stimuli like the white materials (R channel of orange materials). Second the black materials (B channel) display more constancy other

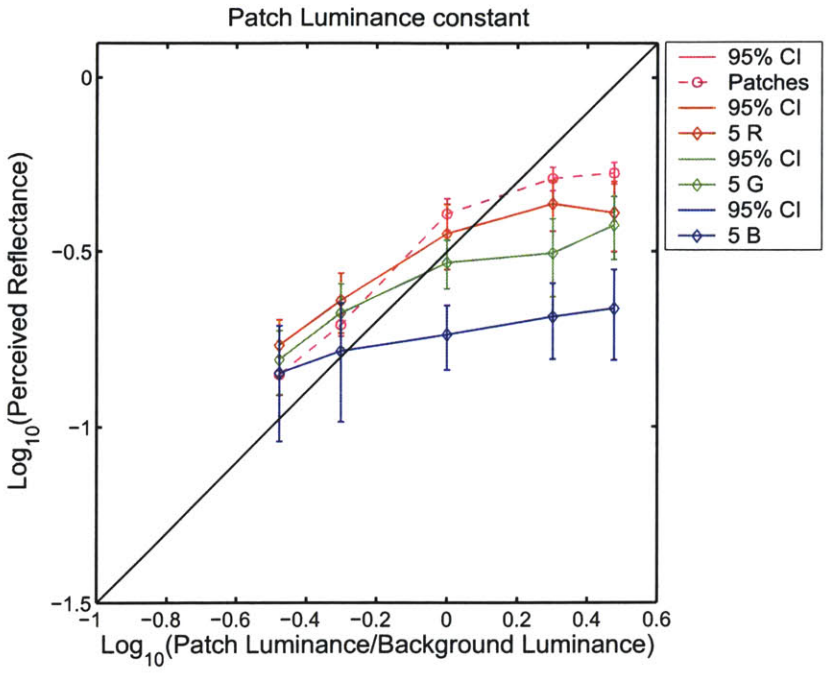
kinds of materials. Moreover the complex visual stimuli are rated in the order of the true reflectances ($R > G > B$).

We can make sense of these findings by appealing to Gilchrist et al's [23] notion of a *framework*. A framework comprises a set of surfaces that belong to the same region of illumination. For our experimental setup (Figure 5-2) the observers views two distinct frameworks - the Munsell box and the LCD. The brightest region in each framework provides an anchor for a surface in the framework (by the highest luminance rule).This is known as local anchoring. Similarly the brightest region in the entire visual field provides the global anchor. The perceived reflectance of any surface in the room will be a compromise between what is predicted by local anchoring and global anchoring.

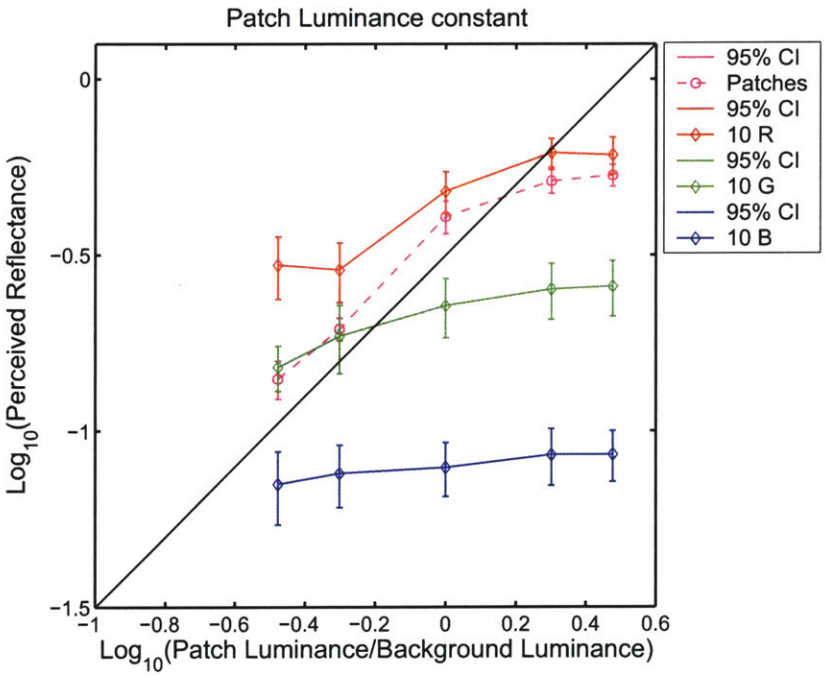
The perceived reflectance of the gray patch as predicted by the local anchoring (if the observers only viewed the LCD) would depend only on the luminance ratio. The reflectance predicted by the global anchor (the brightest region in the room, in our case the white standard patch in the Munsell box, Figure 5-3) depends on the patch luminance. The results in Figures 5-15 to 5-17 are somewhere in between these two predictions. In Figure 5-17 the reflectance ratings change when the luminance ratio is fixed, because the global anchor is important. For Figures 5-15 and 5-15 local anchoring seems to dominate so the reflectance ratings for the patch have a slope close to 1.

For complex visual stimuli like our material images we believe that in addition to local and global anchoring effects, there is a tendency to *self-anchor*. A complex black surface looks dark no matter how much we increase the illumination on the surface. A complex white surface look white even when we reduce the illumination on it. We made similar observations in Chapter 1, in our anti-Gelb experiments. The results in Figures 5-15 to 5-17 indicate self-anchoring is more pronounced for black materials than for the white materials.

Thus, our results with gray patches conform with the existing accounts of lightness perception. They differ from our results with complex stimuli (Experiment IIA) in that observers display greater lightness constancy for images of real world materials. The differences seem to be because real world surfaces have texture and the visual system might use textural cues to judge lightness in addition to cues like mean luminance, surround luminance etc.

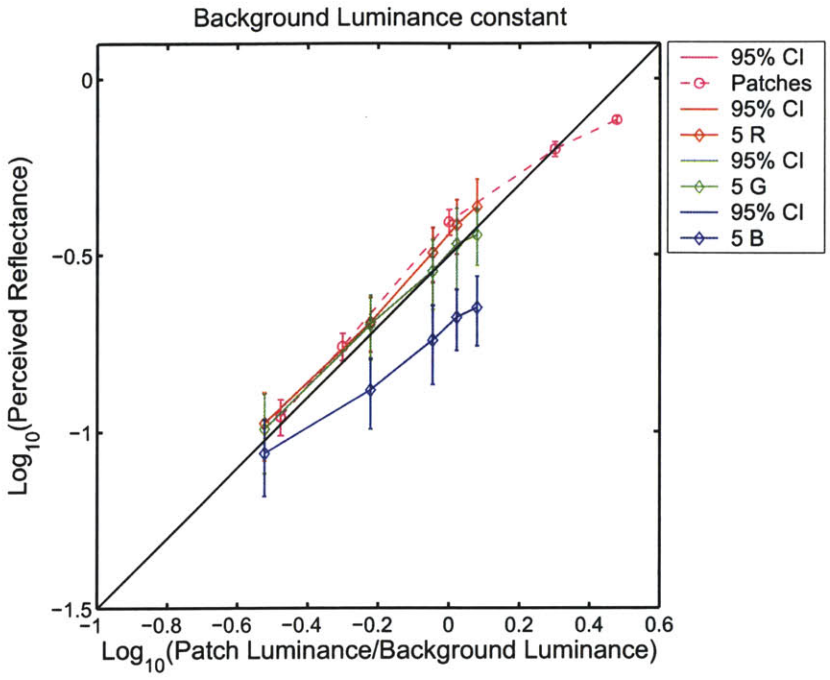


(a)

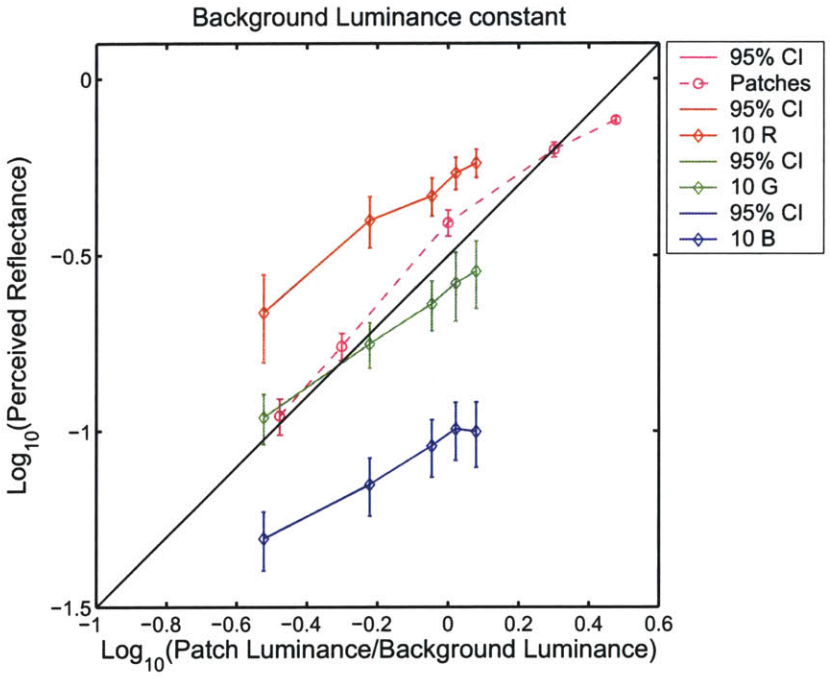


(b)

Figure 5-15: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Patch Luminance/Background Luminance), patch luminance is constant** (a) Comparison with results from Experiment IIA for Material 5 and (b) for Material 10. If observers follow the ratio rule all observations would lie on a line parallel to the black line. The data is pooled across all observers and all lights.

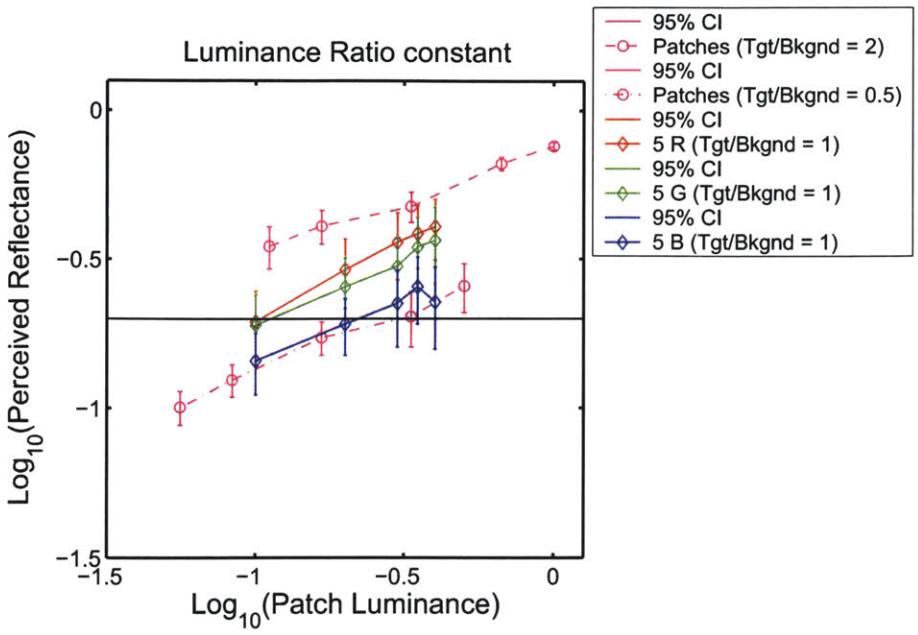


(a)

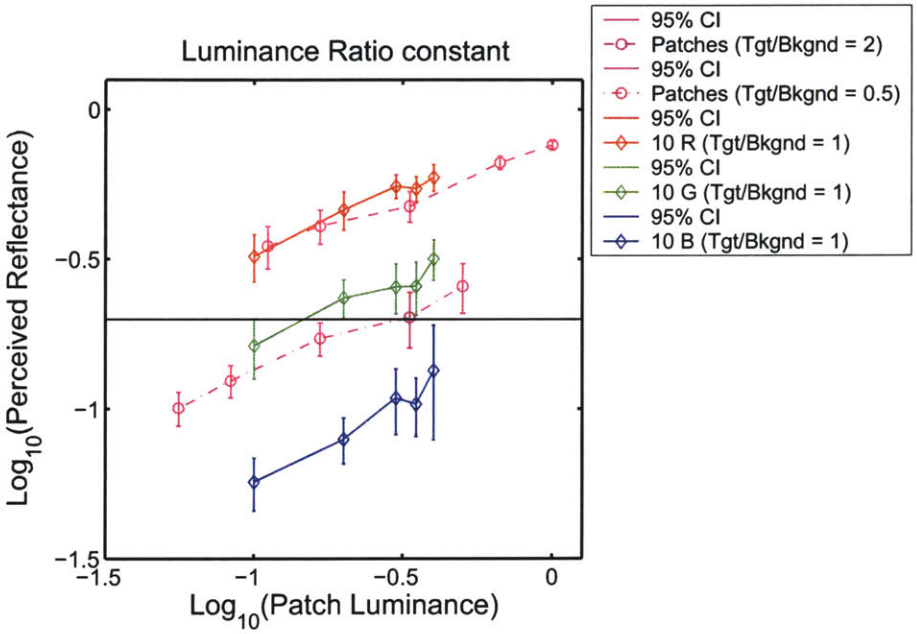


(b)

Figure 5-16: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Patch Luminance/Background Luminance), background luminance is constant** (a) Comparison with results from Experiment IIA for Material 5 and (b) for Material 10. If observers follow the ratio rule all observations would lie on a line parallel to the black line. The data is pooled across all observers and all lights.



(a)



(b)

Figure 5-17: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Patch Luminance), luminance ratio is constant** (a) Comparison with results from Experiment IIA for Material 5 and (b) for Material 10. If observers follow the ratio rule all observations would lie on a line parallel to the black line. The data is pooled across all observers and all lights.

5.8 Experiment IIC

5.8.1 Motivation

In experiment IIA we observed that the background luminance and the mean image luminance affect the perceived reflectance of the material images. However, we also noted that for a given material, the observers consistently rated the R channel higher than the G and G higher than the B. This would suggest that while observers are not perfectly lightness constant, for identical conditions, they can distinguish materials of distinct reflectance values. To test this hypothesis, in this experiment we kept the background and mean image luminance constant and showed observers images of different materials. All channels (R,G, and B) of orange materials and the grayscale images of the non-orange materials were used. We wanted to know if observers rated the images in the order of physical reflectance and how close their reflectance estimates were to the ground truth.

5.8.2 Procedure

Observers viewed images of materials, and indicated ratings for reflectance. The mean luminance of the image and the luminance of the background were both held fixed at 0.33. Twelve observers participated in this experiment. They were divided into three groups, four observers in each. Observers in each group viewed a different set of materials. Four orange materials and three black-white materials under three lights were used for Groups 1 and 3. Three orange materials and four black-white materials under three lights were used in Group 2. For each material, 3 repetitions were run for each lighting condition. For orange materials, 3 repetitions were run per lighting condition per channel. The order of images was randomized. The experiment lasted 30 minutes. In addition to the usual channels R, G and B, for each orange material, the R2B and B2R images for each lighting condition were also included in the experiment.

5.8.3 Results

In this experiment, mean image luminance and the background luminance are both held fixed at 0.33. Subjects view the images of various materials - orange, white and black - under three lighting conditions. For orange materials, the manipulated images R2B and B2R are also included. We do so to test if manipulating statistics leads to a change in perception. Figures 5-18, 5-19 and 5-20 plot the perceived reflectance versus ground truth for each observer in each group. If observers were perfectly lightness constant, their responses would lie along the line with slope 1. If observers display zero constancy their responses would be the same for all materials. The plots show that our observers lie somewhere between these two extremes.

To analyze the performance of the observers further in Figure 5-21 we plot the absolute value of the difference between perceived reflectance and ground truth for each (material, channel, light, observer) combination for each group. In Figure 5-19 one can easily see how observers' responses deviate from the ground truth for each material. For Groups 1 and 3 (Figure 5-23) observers agree with each other and make similar mistakes on the same materials. For Group 2 (Figure 5-22) however, observers display individual differences in their errors. Interestingly, we find that the error depends on the material i.e. some materials are harder to judge than others. Figure 5-21 through 5-23 show the analysis for Light 1. More plots for Light 2 and 3 are in Appendix B.

In Figure 5-24 the absolute error is plotted against the reflectance of the material for each light and each group. From the plots, we observe that the error does not seem to be related to the reflectance of the material.In other words, the deviation of perceived reflectance from the ground truth, does not depend on the actual reflectance (ground truth) of the material.

Finally, we examine the effect of manipulating image statistics. In Figures 5-26

and 5-28, the mean response for each version (R,B2R,G,R2B,B) for each orange material is graphed as a bar plot for each group in each light. These plots allows us to examine the success of image statistics at changing the perception of a material. Figures 5-25 and 5-27 show example stimuli. Image R and B2R have identical histograms and filter output histograms. Therefore according to our chosen set of statistics they are indistinguishable. If these statistics capture anything of perceptual relevance then images R and B2R should be rated identically by all observers. The same reasoning holds for the B and R2B images. From the plots (more plots in Appendix B) we find that for nearly all materials and under all lights, both the (R,B2R) and (B,R2B) mean response pairs are within 2 standard error bars of each other. This is a very satisfying result as it confirms that our chosen statistics capture perceptually relevant image information.

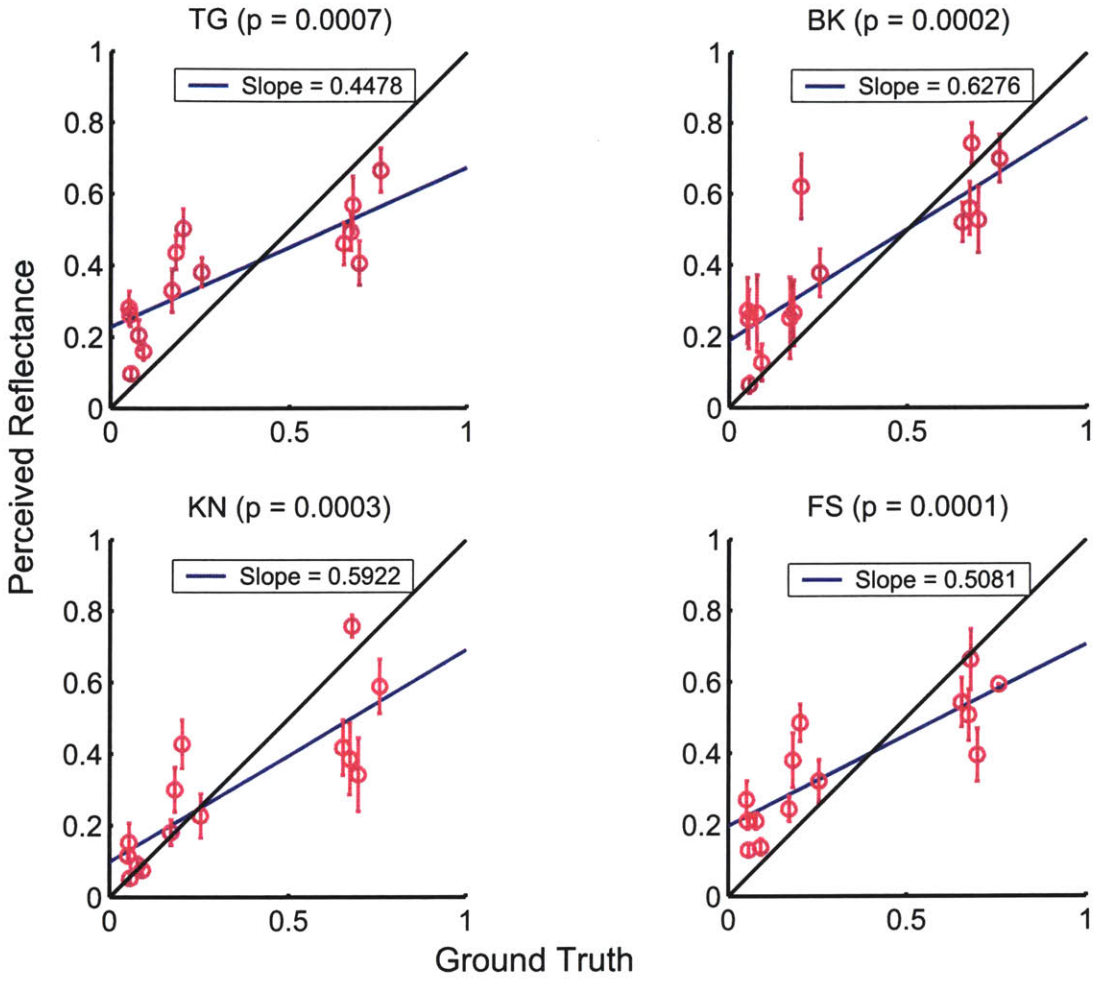


Figure 5-18: **Perceived reflectance vs Ground Truth (Group 1).** The mean responses (pooled over all lighting conditions) are plotted against the ground truth for each observer. Errorbars indicate the 95% confidence intervals. The responses of a veridical observer would lie along the black line with slope = 1 . The blue line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers the slope of the fit is significantly different from 0 and 1.

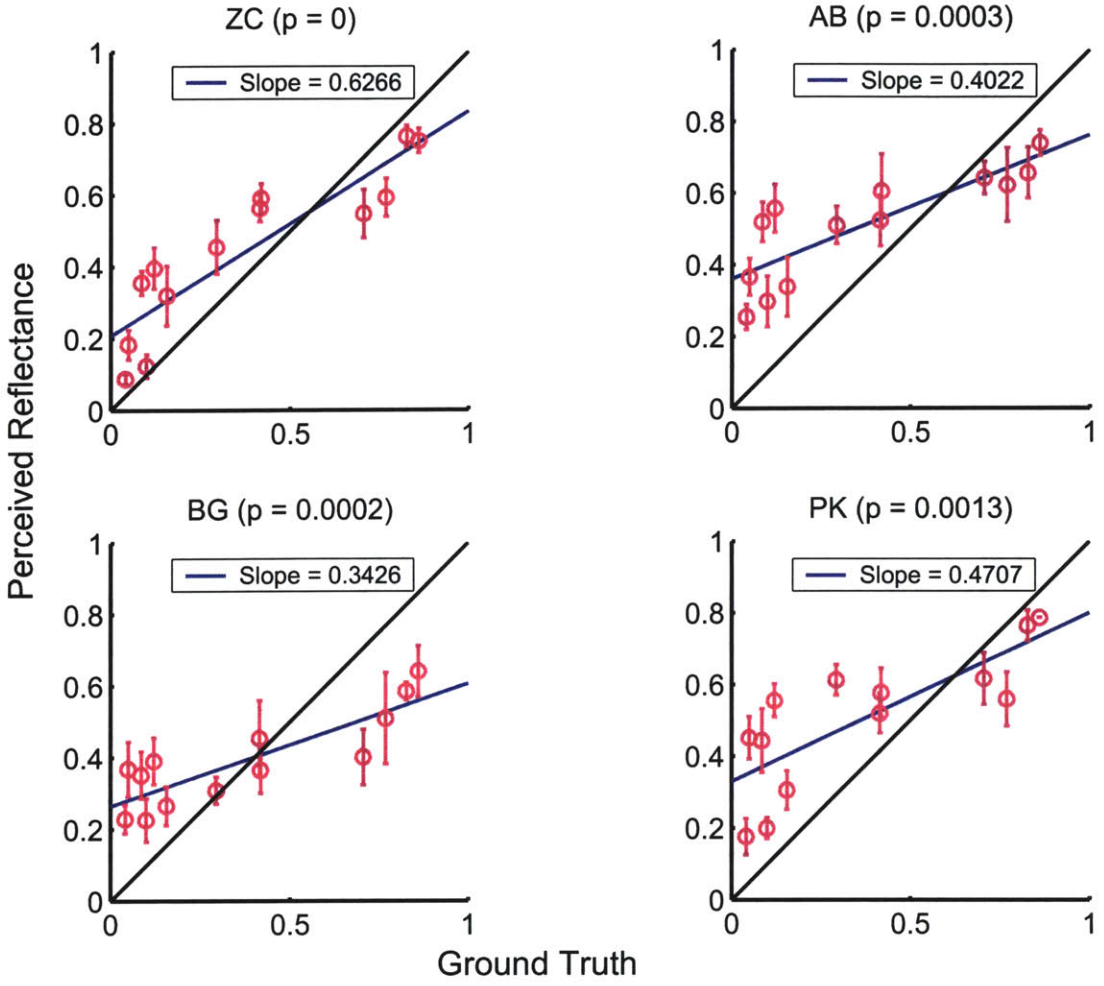


Figure 5-19: **Perceived reflectance vs Ground Truth (Group 2)**. The mean responses (pooled over all lighting conditions) are plotted against the ground truth for each observer. Errorbars indicate the 95% confidence intervals. The responses of a veridical observer would lie along the black line with slope = 1 . The blue line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers the slope of the fit is significantly different from 0 and 1.

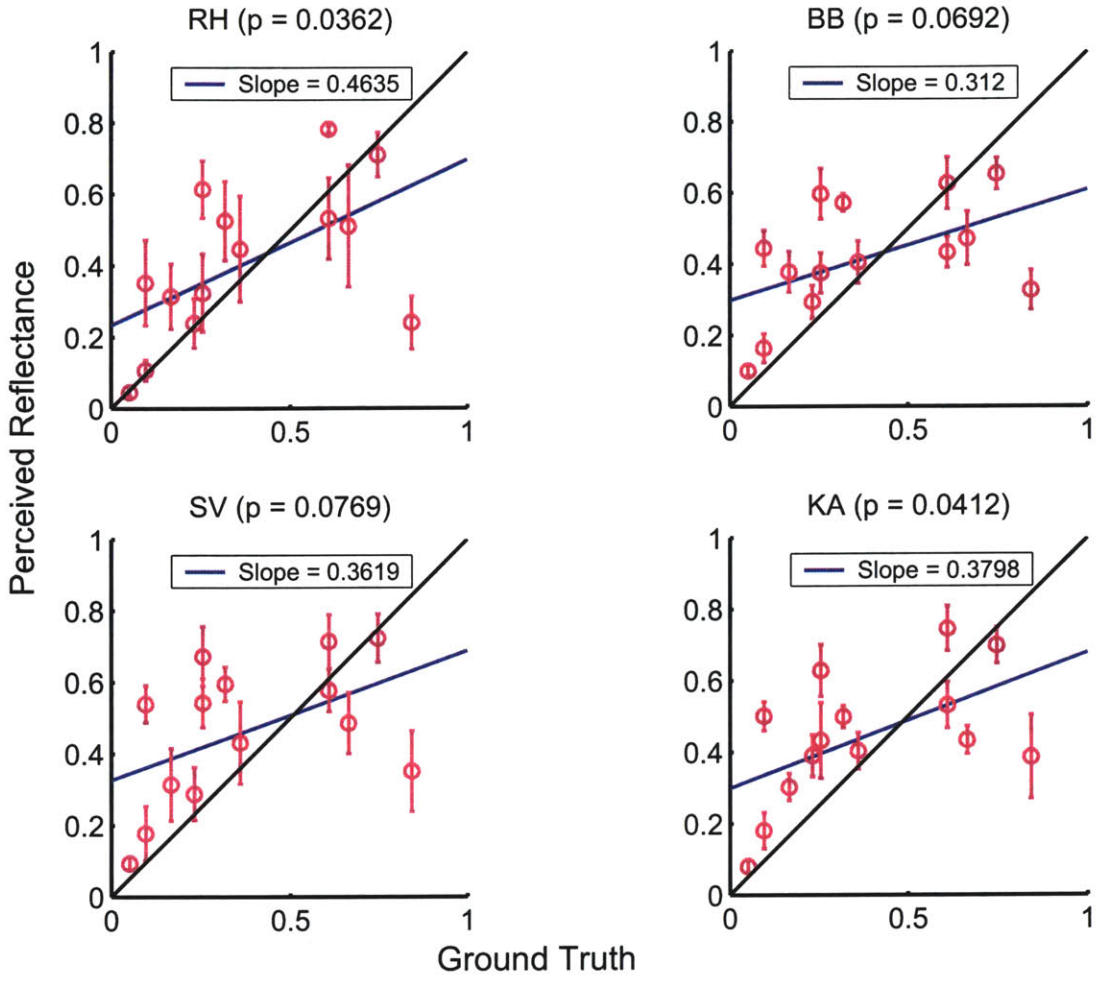


Figure 5-20: **Perceived reflectance vs Ground Truth (Group 3)**. The mean responses (pooled over all lighting conditions) are plotted against the ground truth for each observer. Errorbars indicate the 95% confidence intervals. The responses of a veridical observer would lie along the black line with slope = 1 . The blue line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For two observers, RH and KA, the slope of the fit is significantly different from 0 and 1. For observers BB and SV, the slope is not significant.

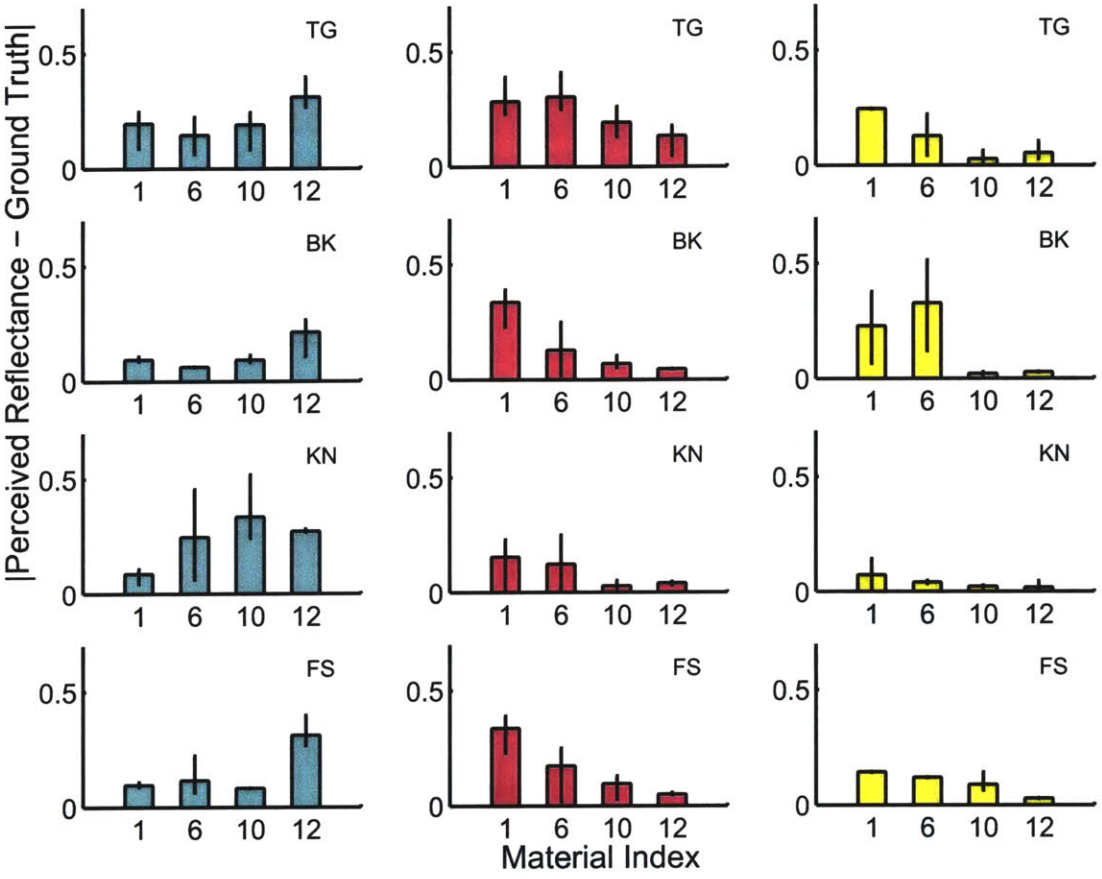


Figure 5-21: **|Perceived reflectance - Ground Truth| vs Material Index (Group 1, Light 1)**. The mean absolute difference between ground truth and perceived reflectance is plotted against the material for each observer. Errorbars indicate the range. Each row corresponds to a different observer and each column to the R (cyan), G (magenta) or B (yellow) channels for each material. Note how most observers agree with each other and how some materials are harder to judge than others.

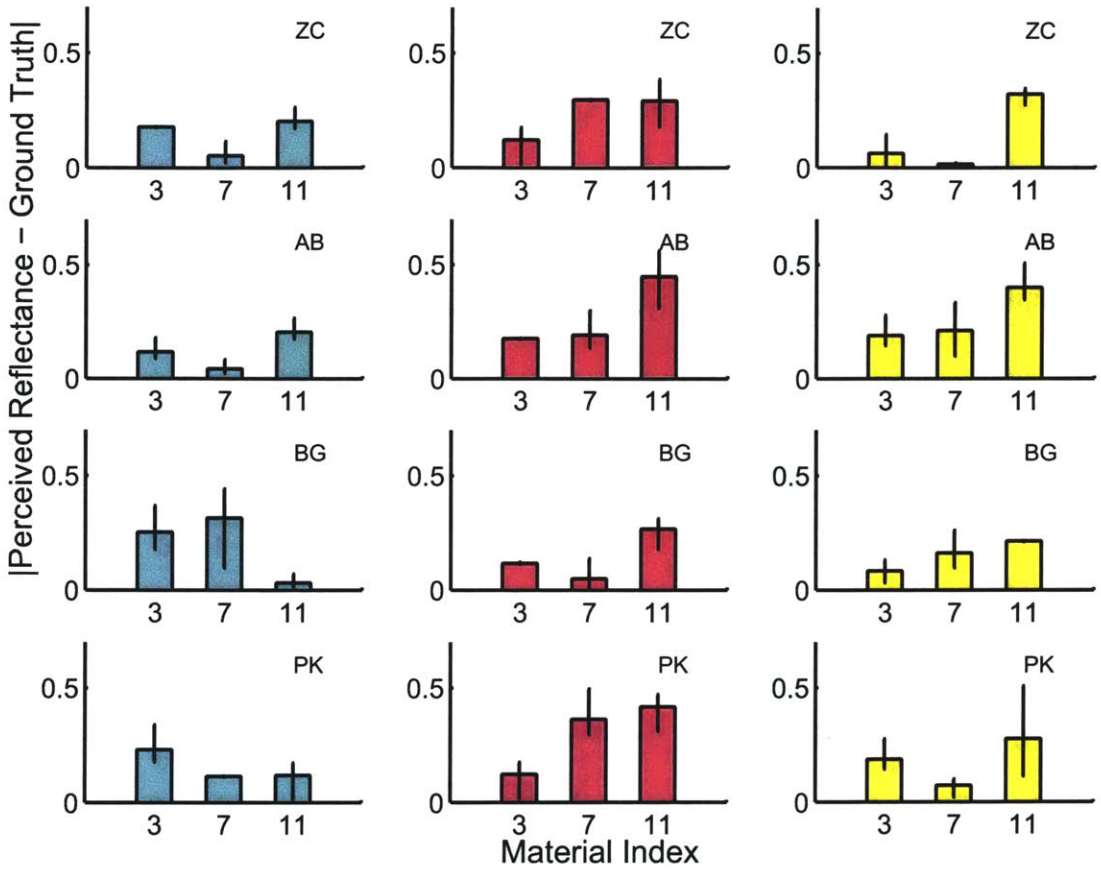


Figure 5-22: **|Perceived reflectance - Ground Truth| vs Material Index (Group 2, Light 1)**

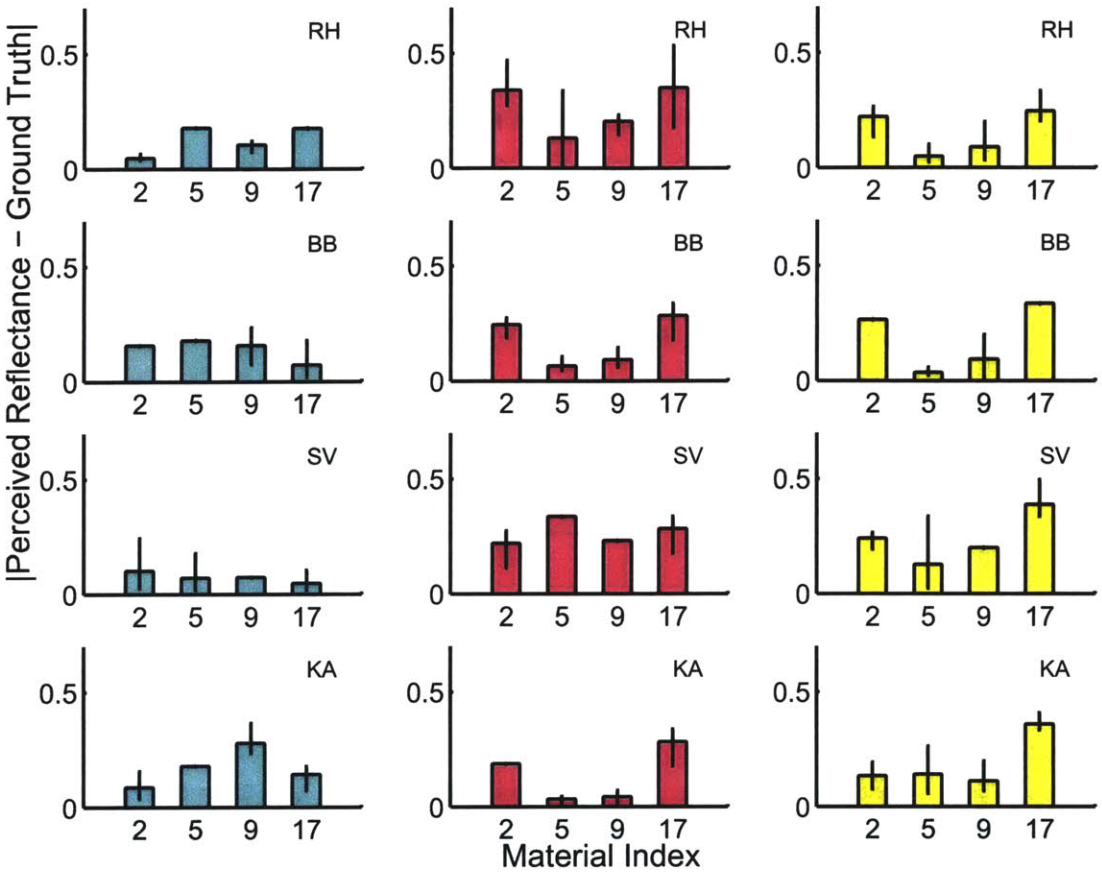


Figure 5-23: **|Perceived reflectance - Ground Truth| vs Material Index (Group 3, Light 1)**

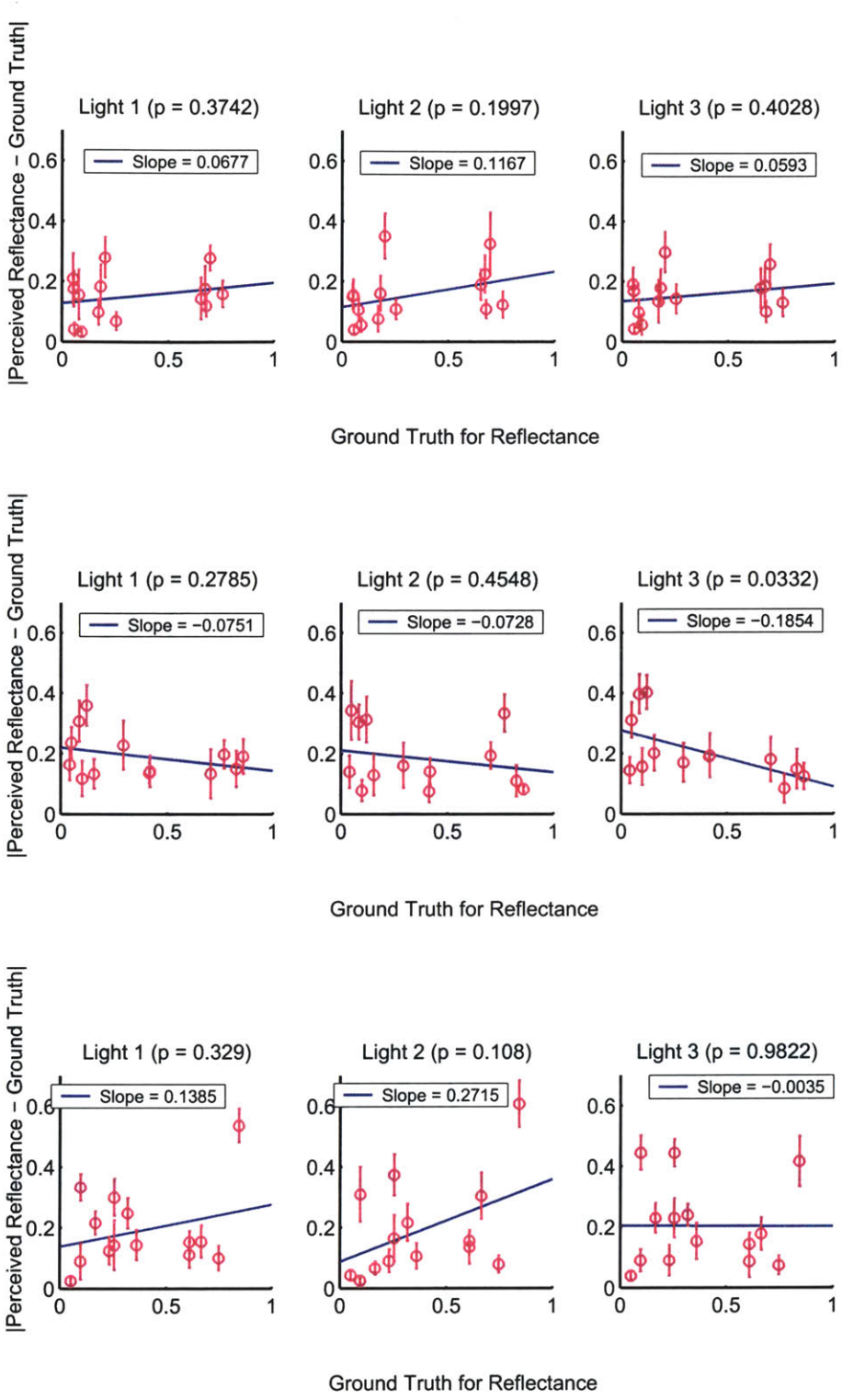


Figure 5-24: **|Perceived reflectance - Ground Truth| vs Ground Truth** (Top row) Group 1, all lights (Middle row) Group 2, all lights (Bottom row) Group 3, all lights

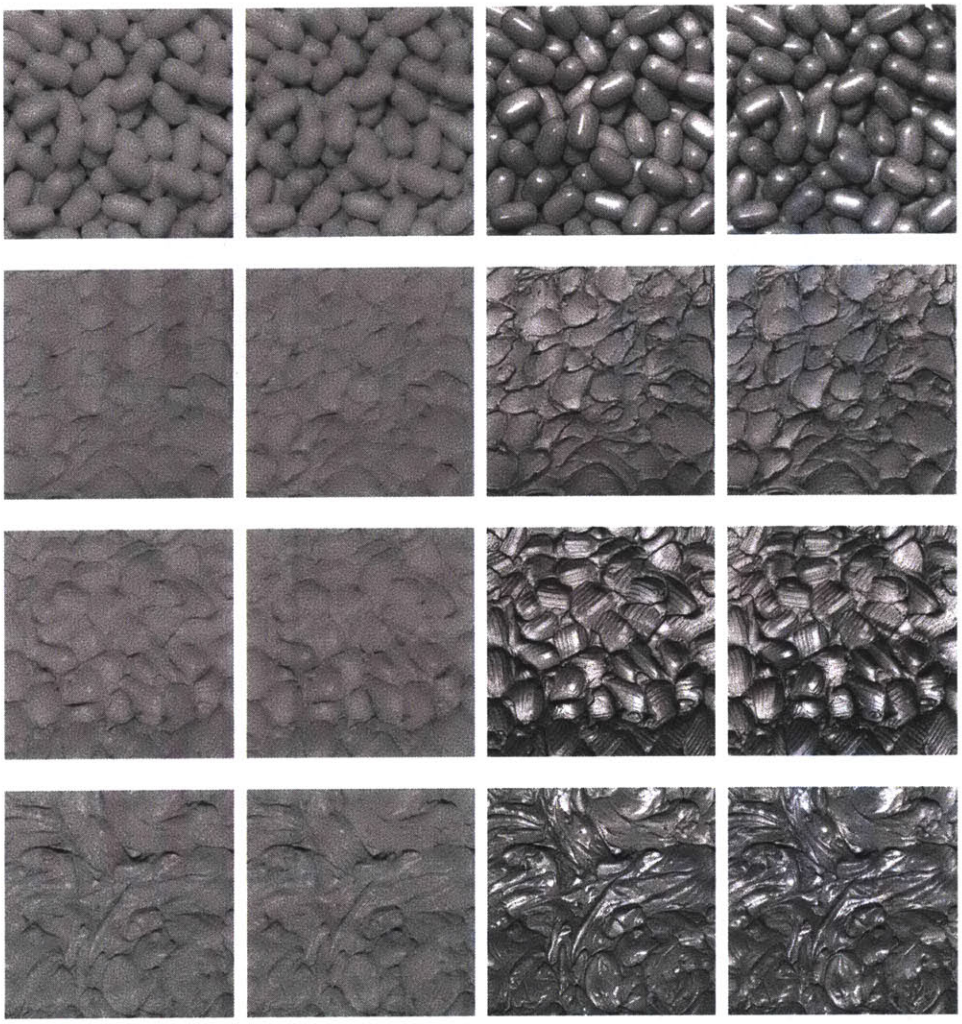


Figure 5-25: Stimuli for Experiment IIC, Group 1 (Light 1) : The columns are the R, R2B, B2R and B images of all materials in Group 1. (Top row) Material 1 (Second row) Material 6 (Third row) Material 10 (Fourth row) Material 12

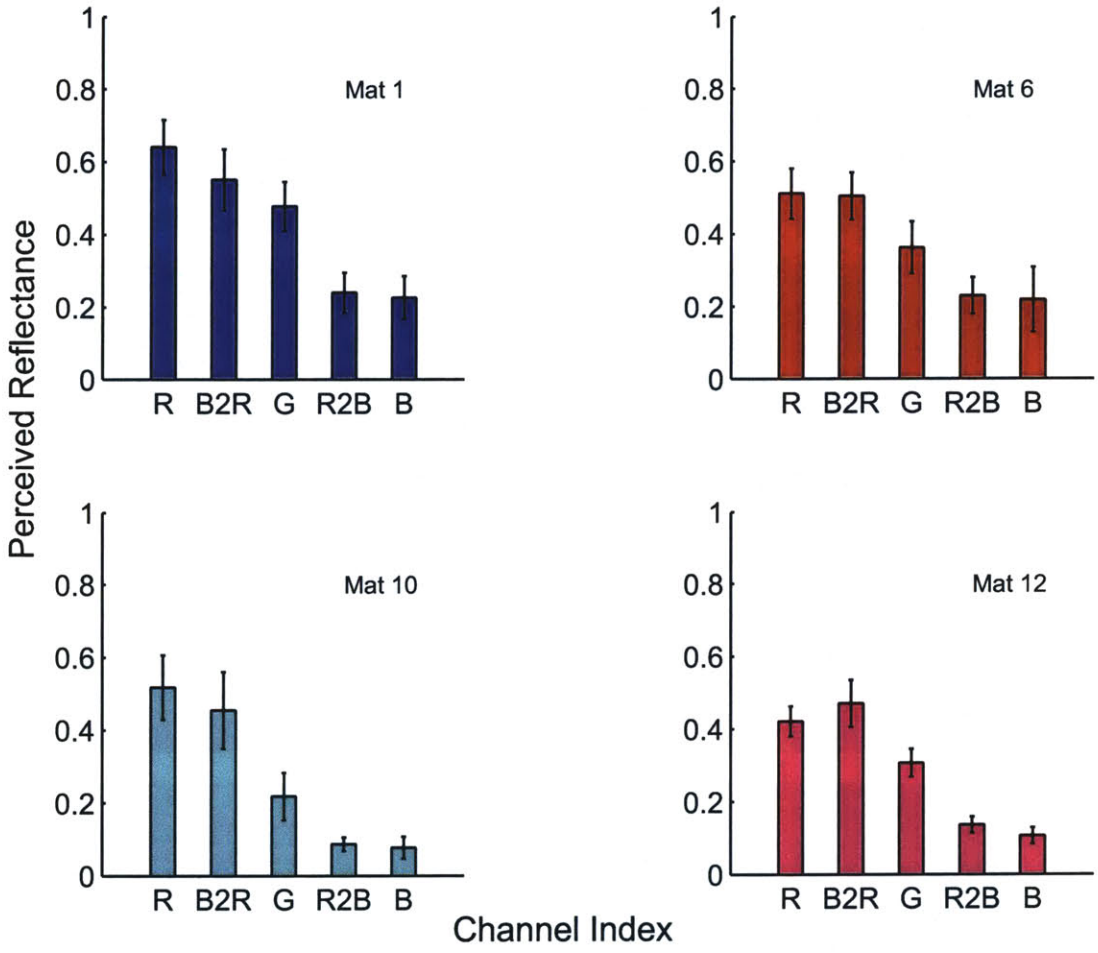


Figure 5-26: **Perceived Reflectance vs Channel Index (Group 1, Light 1)** The mean response pooled across subjects is plotted against the channel index for each material. Error bars indicated 95% confidence intervals. The success of manipulated images $R2B$ and $B2R$ varies depending on the material.

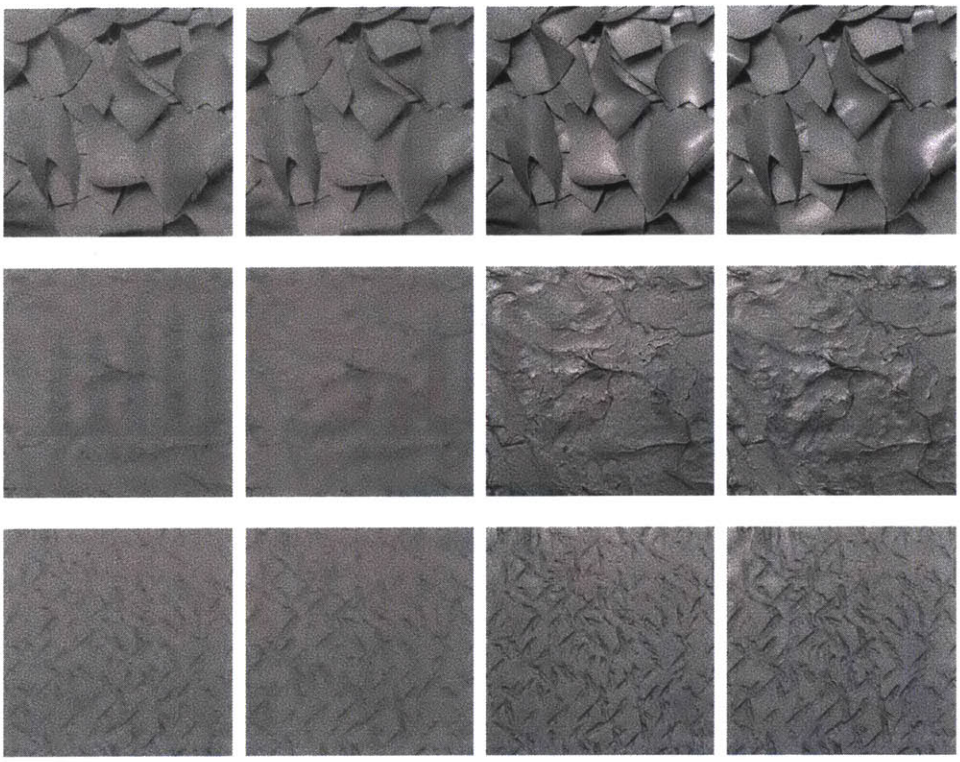
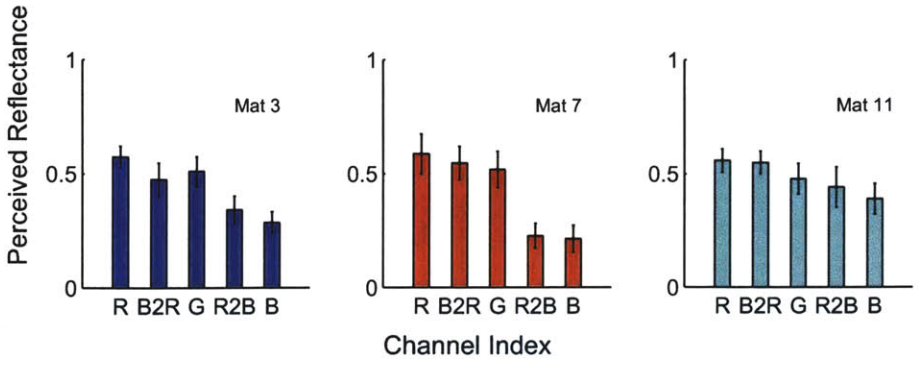
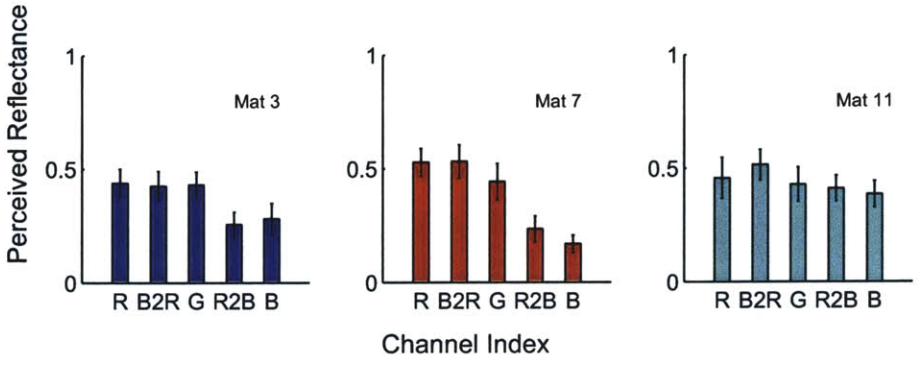


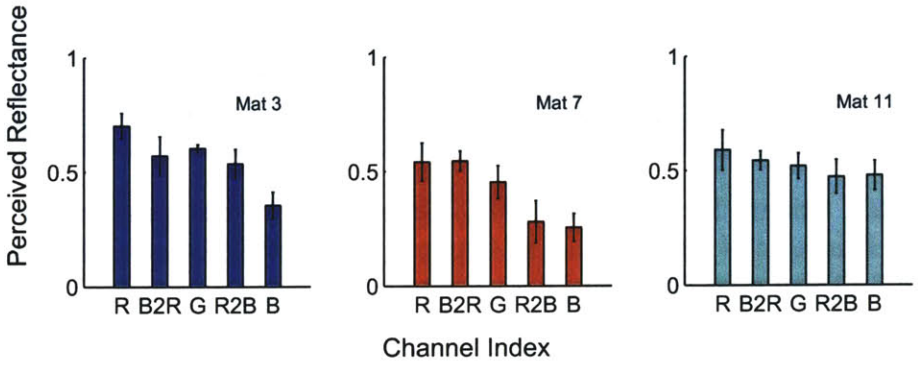
Figure 5-27: Stimuli for Experiment IIC, Group 2 : The columns are the R, R2B, B2R and B images of all materials in Group 2. (Top row) Material 3 (Second row) Material 7 (Third row) Material 11



(a) Light 1



(b) Light 2



(c) Light 3

Figure 5-28: **Perceived Reflectance vs Channel Index (Group 2)** The mean response pooled across subjects is plotted against the channel index for each material. Error bars indicated 95% confidence intervals. The success of manipulated images $R2B$ and $B2R$ varies depending on the material.

5.9 Comparison of Experiment IIc responses with Statistics

The results of experiments IIC, particularly the result that manipulated images R2B and B2R were rated nearly the same as the B and R images respectively, suggest that our choice of image statistics is a good one. The statistics agree to a great extent with perceptual judgments of reflectance. In Section 5.5.3 we observed that the performance of a simple linear regression classifier using three image statistics could rival human performance. We would like to compare the performance of a learning method that uses image statistics to that of the observers in Experiment IIC. We want to investigate how good the statistics are at estimating reflectance as opposed to classifying reflectance.

Unlike Experiment I, where the observers had a training phase, the comparison between a learning method that employs image statistics and humans is unequal in this case because the observers have no access to ground truth. For any learning algorithm to succeed, training data is necessary. Hence, we will do the following - we train an estimator based on image statistics on images in Group X (X = 1,2 or 3) with ground truth. Then we test the estimator on the remaining groups and compare the performance and errors with observers.

An epsilon Support Vector Regression technique is used with a linear kernel [8]. The estimator uses four statistics - the standard deviation and $(90^{th} - 10^{th})$ percentile of the gaussian center surround filter output and Sobel filter output at the finest scale. The parameter ϵ is set to 0.1 and the penalty parameter C is chosen by a five-fold cross validation on the training set. The performance on the training set (Group 1) is shown in Figure 5-29. We see that the even though the estimator is trained on the ground truth, the performance is far from perfect. Comparing the errors made by the estimator on the training set with those made by observers is more illustrative (refer Figure 5-30, error bars for observers are the 95% confidence intervals, the error

bar for the estimator is the range, x axis is the material index). We observe that the errors made by the estimator and the observers are indistinguishable. This is interesting because the estimator trains on the ground truth and has no clue how humans perform the same task. From Figure 5-30 we see that it is impossible to tell apart the performance of the machine and an average human observer in this experiment.

The performance of the estimator on test sets (Group 2 and Group 3) is plotted in Figures 5-31 and 5-32. The overlaps in errors of the estimator and observers for Group 2 and Group 3 is significant but not perfect. Figures 5-33 through 5-36 graph the performance for the case when the estimator is trained on Group 2 images and tested on Groups 3 and 1. Figures 5-37 through 5-40 graph the performance for the case when the estimator is trained on Group 3 images and tested on Groups 1 and 2. Using non-linear kernels like polynomial or radial basis function with the ϵ-SVR technique either leads to performance identical to the linear case or overfitting.

Thus, even with four features and a reasonably simple learning method, the learning algorithm rivals human performance at estimating reflectance. Moreover, they make the same mistakes as the observers. This leads us to conclude that simple image statistics like moments and percentiles of image intensities and filter outputs capture perceptually relevant image information.

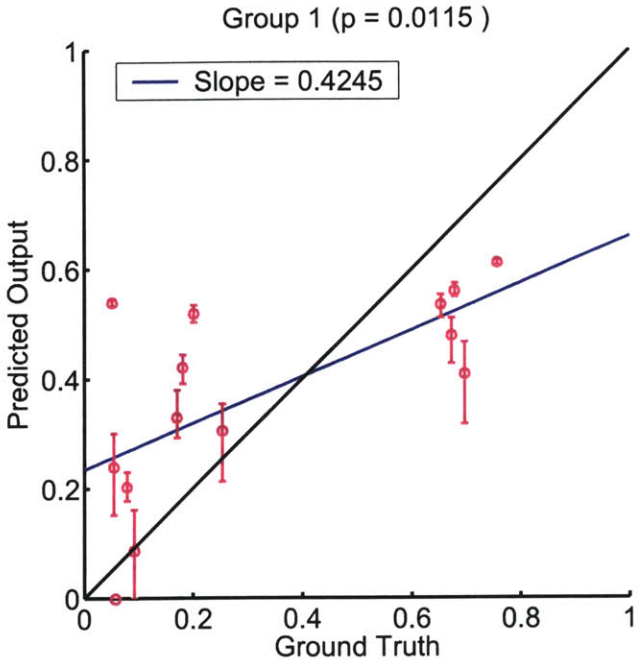
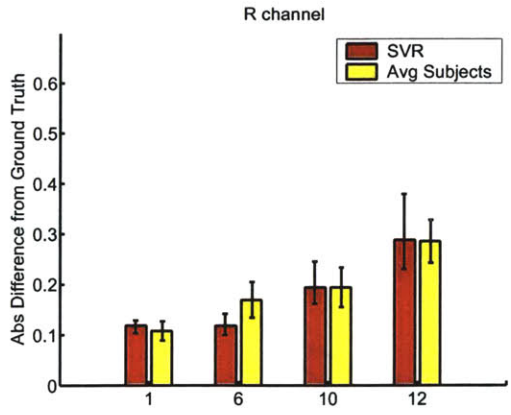
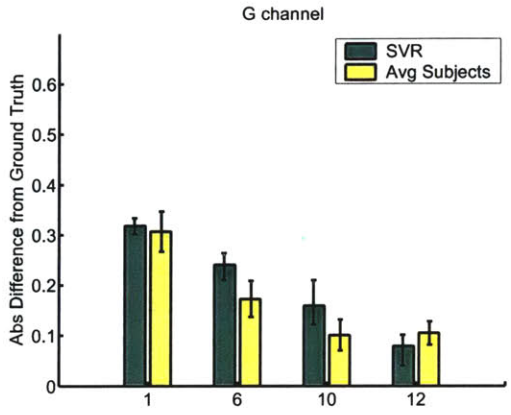


Figure 5-29: **Training Performance ϵ-SVR, Linear Kernel, Set 1 statistics** [1], **train set Group 1** - standard Deviation and $(90^{th} - 10^{th})$ percentile of histograms of gaussian center surround and Sobel filtered images. Group 1 images form the training set. Penalty parameter C is chosen by a five-fold cross validation on the training set. ϵ is set to 0.1.

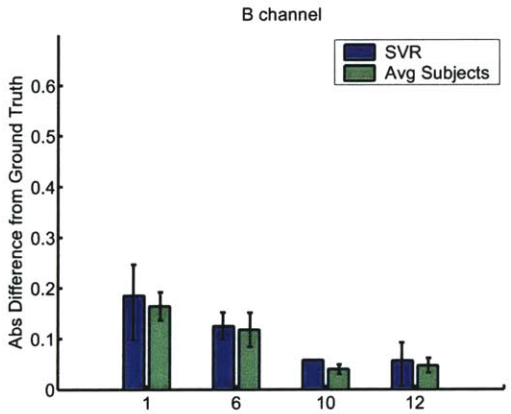
[1]Set 1 statistics are the standard deviation and $(90^{th} - 10^{th})$ percentile of the gaussian center surround filter output and Sobel filter output at the finest scale.



(a)

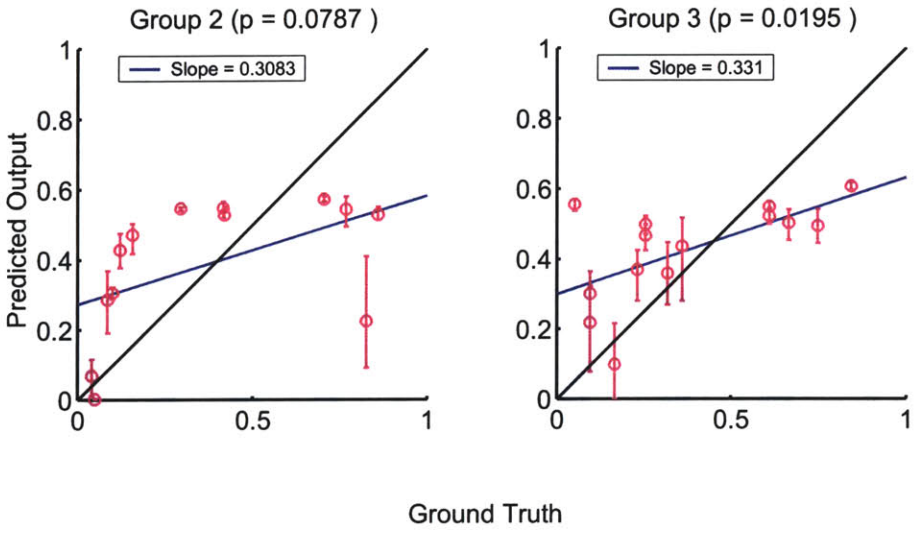


(b)



(c)

Figure 5-30: Comparison of ϵ-SVR with linear kernel, Set 1 statistics with averaged subject performance for training set Group 1



(a)

Figure 5-31: **Test Performance ϵ-SVR, Linear Kernel, Set 1 statistics, train set Group 1** The predicted output is plotted versus ground truth for test groups 2 and 3.

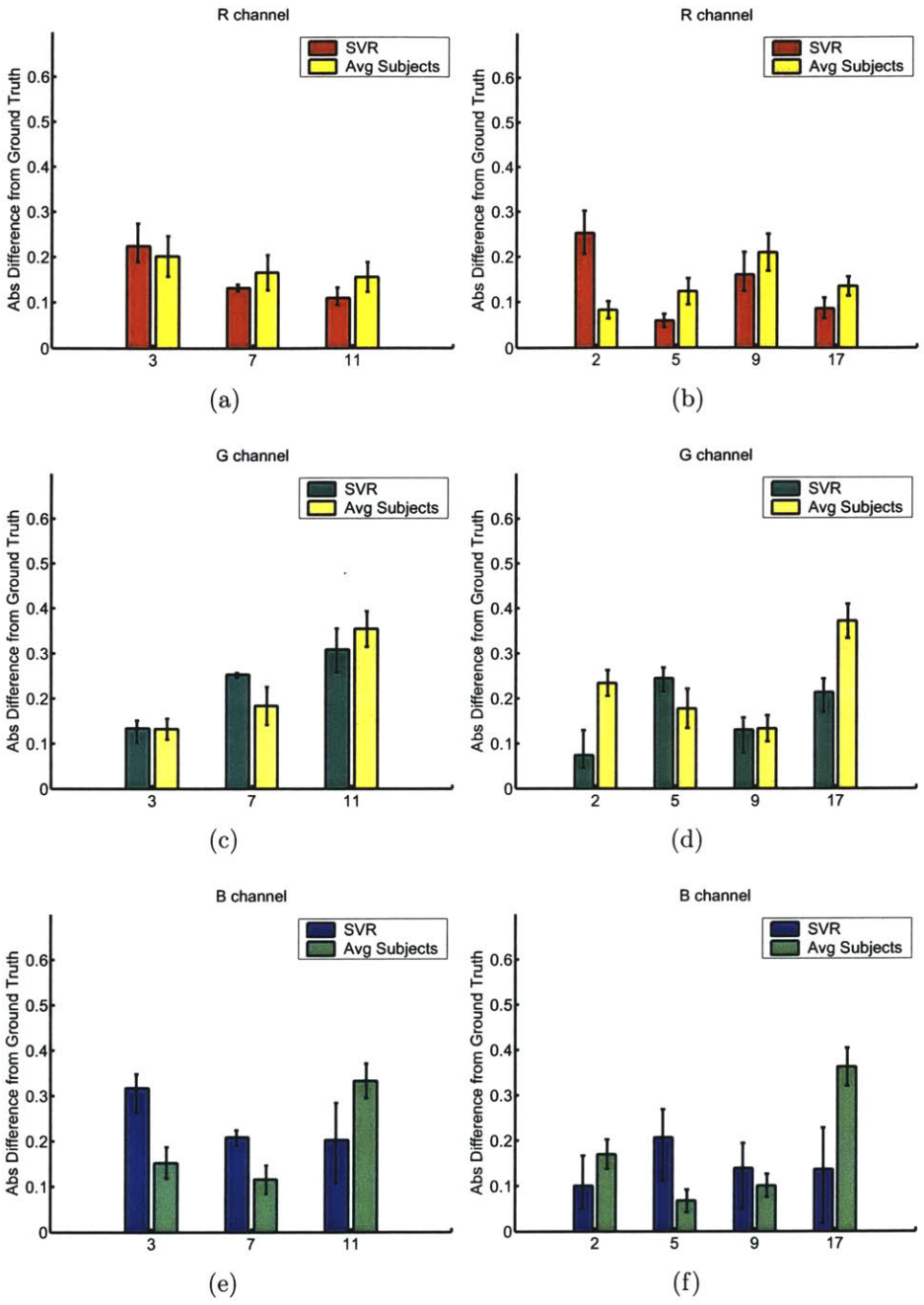


Figure 5-32: Comparison of ϵ-SVR with linear kernel, Set 1 statistics with averaged subject performance for test sets - Group 2 (left column) and Group 3(right column). Each row corresponds to the R, G and B channels of the materials.

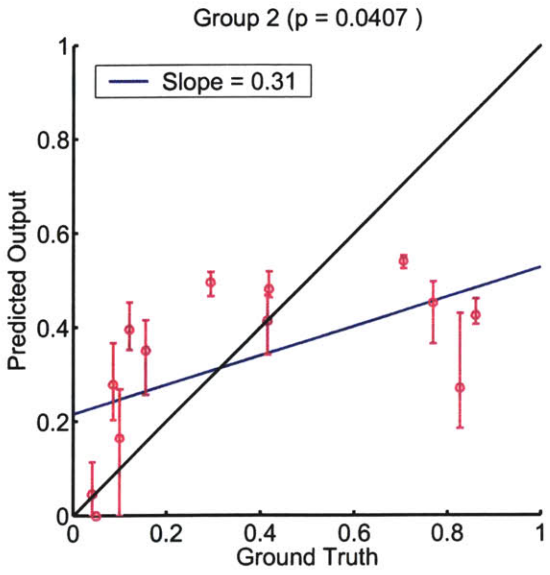
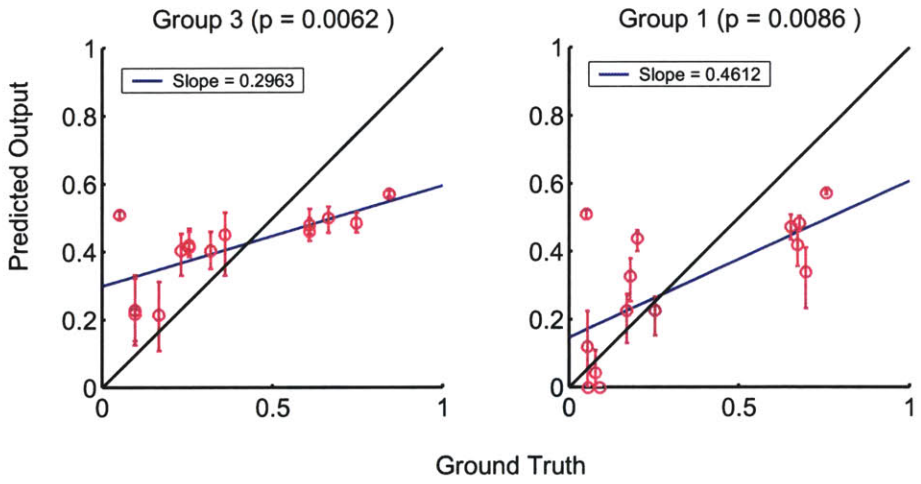
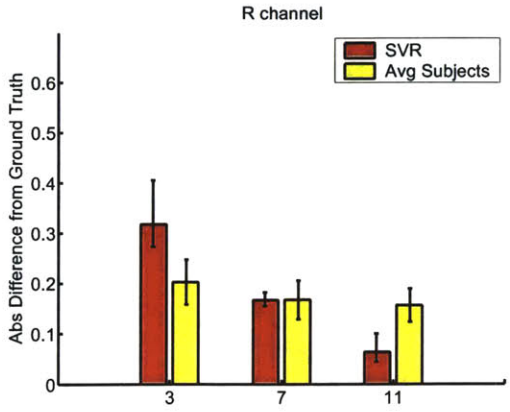


Figure 5-33: **Training Performance ϵ-SVR, Linear Kernel, Set 1 statistics, train set Group 2** - standard Deviation and $(90^{th} - 10^{th})$ percentile of histograms of gaussian center surround and Sobel filtered images. Group 2 images form the training set. Penalty parameter C is chosen by a five-fold cross validation on the training set. ϵ is set to 0.1.

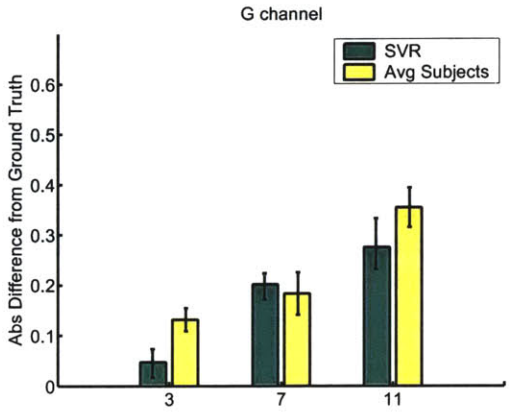


(a)

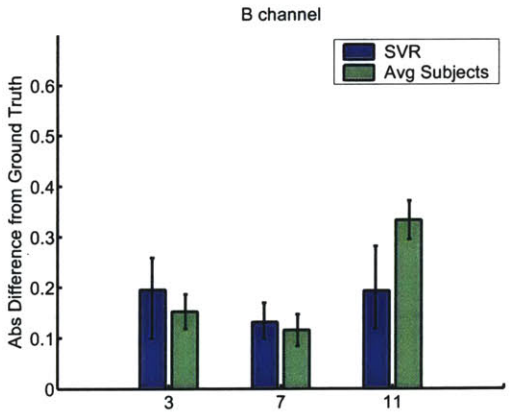
Figure 5-34: **Test Performance ϵ-SVR, Linear Kernel, Set 1 statistics, train set Group 2** The predicted output is plotted versus ground truth for test groups 3 and 1.



(a)



(b)



(c)

Figure 5-35: Comparison of ϵ-SVR with linear kernel, Set 1 statistics with averaged subject performance for training set Group 2

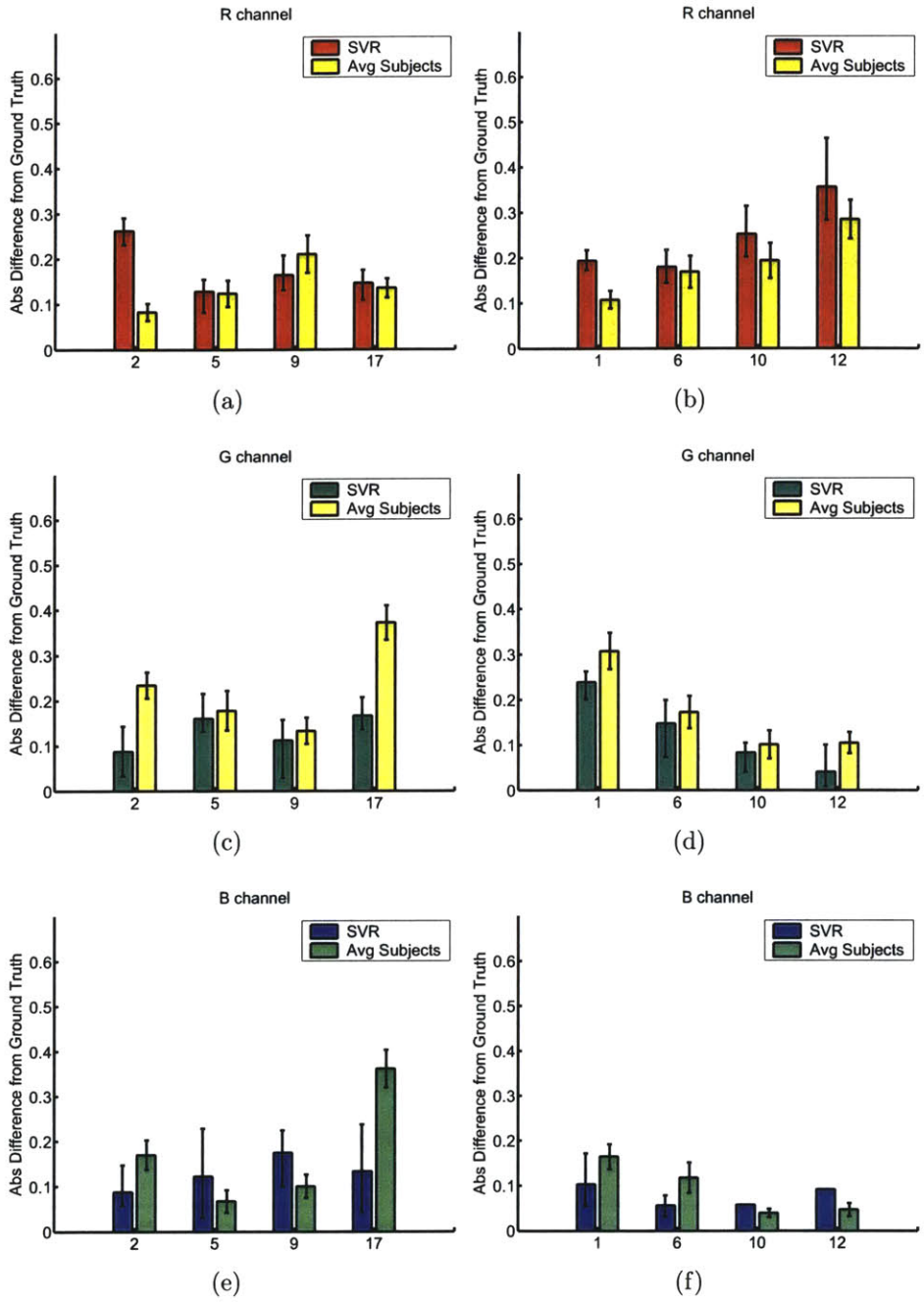


Figure 5-36: Comparison of ϵ-SVR with linear kernel, Set 1 statistics with averaged subject performance for test sets - Group 3 (left column) and Group 1(right column). Each row corresponds to the R, G and B channels of the materials.

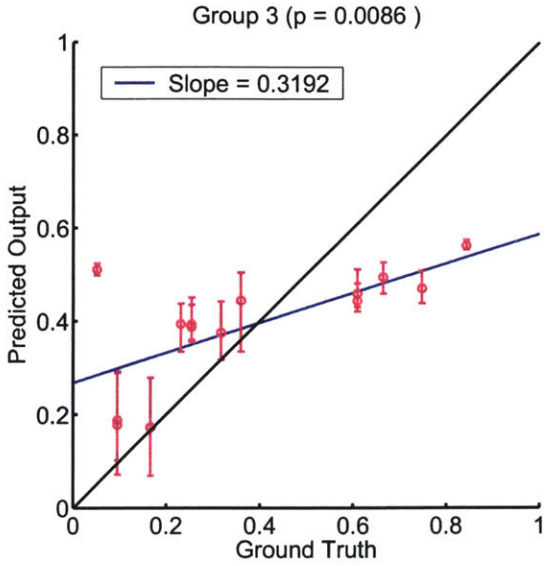
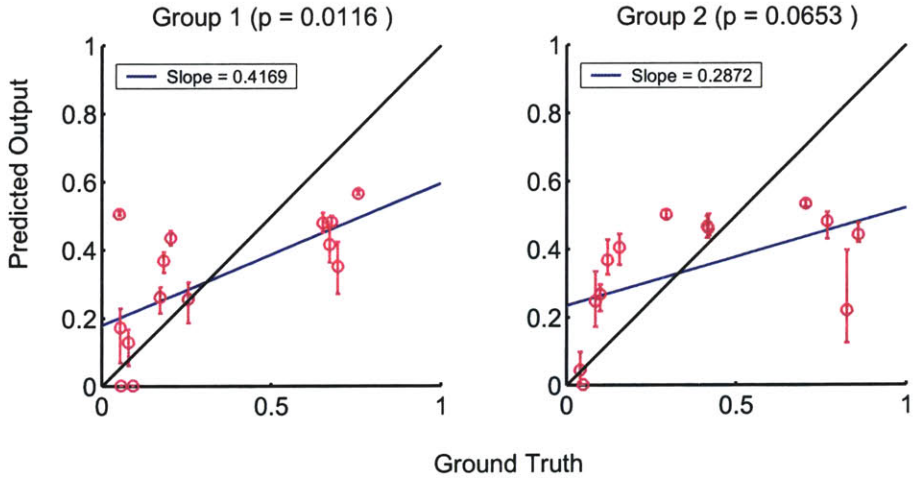
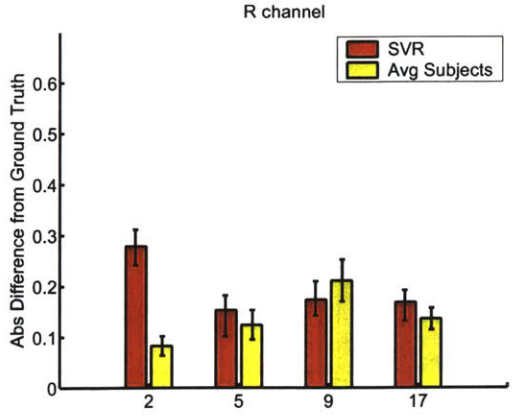


Figure 5-37: **Training Performance ϵ-SVR, Linear Kernel, Set 1 statistics, train set Group 3** - standard Deviation and $(90^{th} - 10^{th})$ percentile of histograms of gaussian center surround and Sobel filtered images. Group 3 images form the training set. Penalty parameter C is chosen by a five-fold cross validation on the training set. ϵ is set to 0.1.

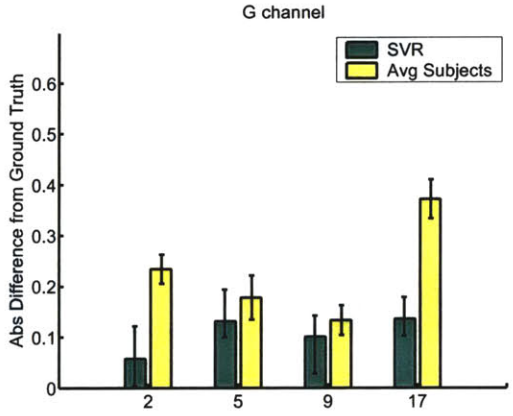


(a)

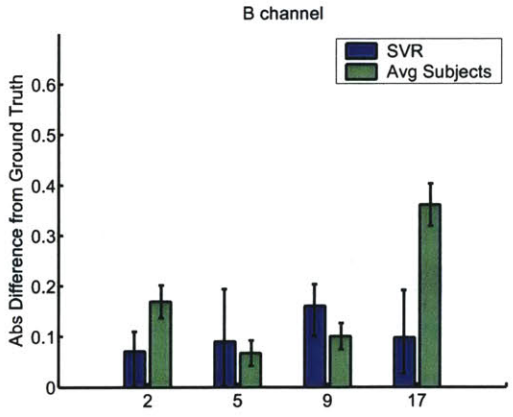
Figure 5-38: **Test Performance ϵ-SVR, Linear Kernel, Set 1 statistics, train set Group 3** The predicted output is plotted versus ground truth for test groups 1 and 2.



(a)



(b)



(c)

Figure 5-39: Comparison of ϵ-SVR with linear kernel, Set 1 statistics with averaged subject performance for training set Group 3

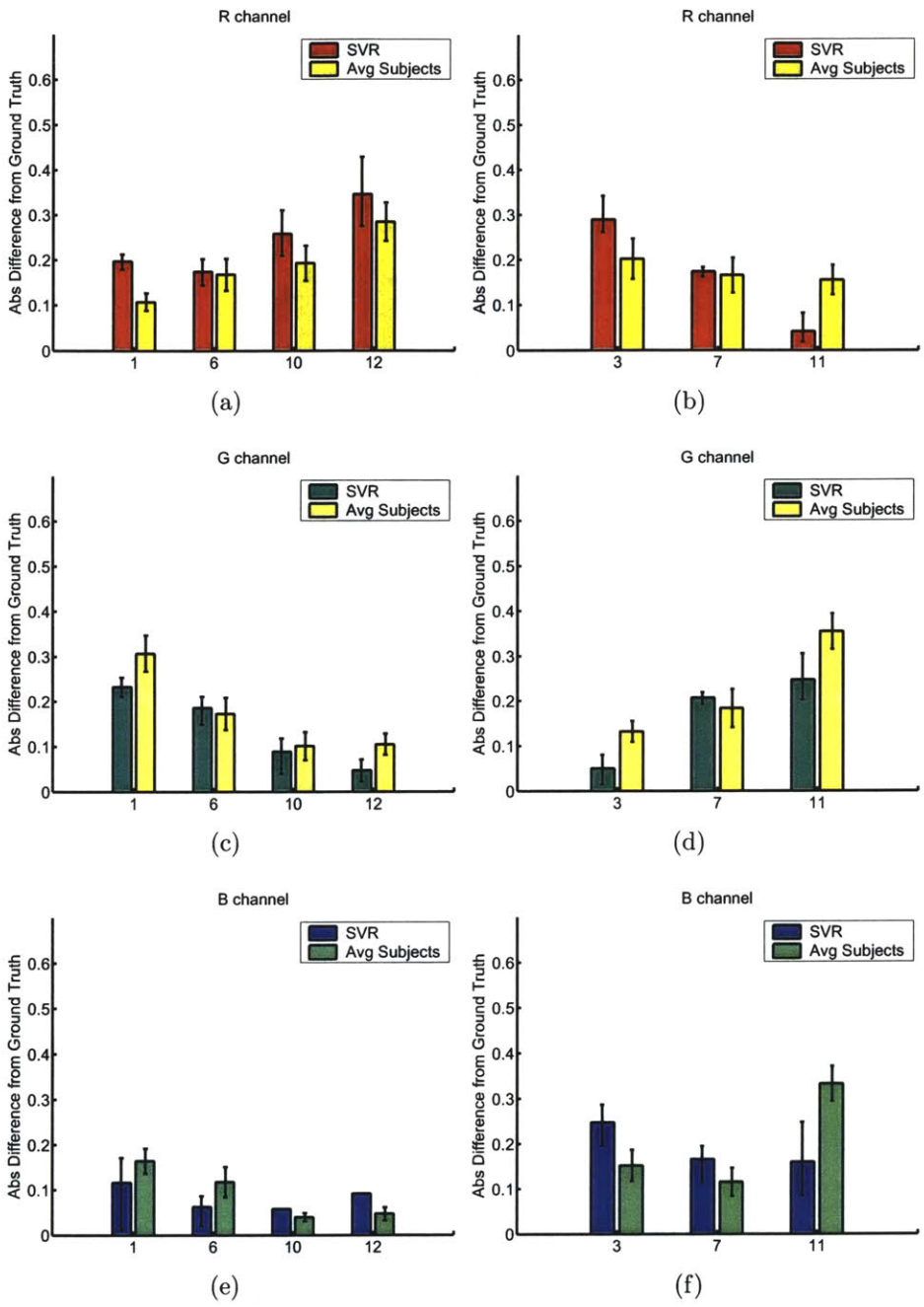


Figure 5-40: Comparison of ϵ-SVR with linear kernel, Set 1 statistics with averaged subject performance for test sets - Group 1 (left column) and Group 2(right column). Each row corresponds to the R, G and B channels of the materials.

Chapter 6

Summary

In this thesis we contribute to progress in the material recognition problem, specifically in surface reflectance recognition. Dror et al [16, 14, 15] formulated the reflectance recognition problem in a statistical framework. They showed statistical regularities in real world illumination lead to informative relationships between the reflectance of a surface and certain statistics measured on an image of the surface. In this work, we cast the reflectance recognition problem in the same framework. We restrict our reflectance space to opaque materials of spatially uniform reflectance properties. We allow materials to possess surface mesostructure. We find that for such materials certain image statistics are diagnostic of the diffuse surface reflectance. We compare the performance of a learning algorithm that uses such statistics as features to human observers in two psychophysical experiments. We find that learning algorithms that employ such image statistics perform similarly to an average human observer.

In Section 6.1 we discuss our contributions in detail and in section 6.2 we outline directions for further research.

6.1 Thesis Contributions

6.1.1 Shadows, Interreflections and Surface Mesostructure

One of the important contributions of this work is that we consider surfaces with mesostructure (medium scale structure that can be resolved by the eye). Interaction of light with such surfaces leads to shadows and interreflections. Estimating the reflectance of such surfaces from a single image under unknown lighting is considerably harder than for smooth surfaces of known geometry. Image based reflection estimation techniques [58, 57, 5, 35, 42, 12] consider similar surfaces, however all of them require additional information in the form of prior knowledge of illumination or geometry or multiple photographs or human interaction.

In contrast Dror et al [16] assume their surfaces are smooth spheres and thereby avoid shadows and interreflections. However they handle complex real world illumination while we assume simple artificial illumination. Also, Dror et al consider both the specular as well as the diffuse components of reflectance while we focus only on the diffuse component.

6.1.2 Image Statistics and Reflectance Estimation

Like Dror et al, we find that moment and percentile statistics of image intensity histograms and histograms of filtered images are diagnostic of surface reflectance. The statistics are not perfectly correlated with the (diffuse) surface reflectance. However they can be combined by a regression algorithm to predict the reflectance of a surface given a single image with unknown surface geometry and illumination. We find that even with a few features (three to four) and a relatively simple learning algorithm, we can estimate the surface reflectance as well as humans.

Clearly our methods in their current form cannot be applied to estimate the ground

truth for reflectance i.e. full BRDF estimation is impossible. However our findings illustrate that even in a severely underconstrained case (single image, complex surface, unknown lighting) it is possible to estimate the reflectance properties just as well as human observers.

6.1.3 Psychophysics with Complex Stimuli

We conducted psychophysical experiments with images of real world surfaces, in contrast to most of the prior work on lightness perception where the stimuli are flat, Lambertian surface patches. Recently some authors [43, 19, 20, 32] have used complex stimuli in their studies. We believe that using stimuli representative of real world conditions aids understanding the workings of the human visual system. Our visual system excels at interpreting natural world scenes. Therefore, it is plausible that the visual system would perform sub-optimally for stimuli that are not representative of the real world.

In our experiments we find that the classic Gelb effect fails for images of real world textured surfaces. Observers can estimate the reflectance of surfaces in the absence of mean luminance information and context. However, observers are not veridical. Our findings suggest that humans use textural cues in addition to mean luminance information and contextual cues in order to make lightness judgements.

6.2 Future Work

6.2.1 Relaxing constraints on illumination and reflectance

In our work, we assume simple artificial illumination and surfaces with spatially uniform reflectance properties. It should be possible to relax these constraints to include real world complex illumination and more challenging surfaces. We know from our

daily visual experience, that humans can estimate reflectance properties under such conditions. Therefore, it is conceivable that there exist informative image statistics that make reflectance estimation under such challenging conditions feasible. Recent work in texture analysis [60, 41] suggests that joint statistics of subband coefficients and outputs of non-linear filters capture perceptually relevant characteristics of textures. As the distinction between arbitrary surfaces and textures is hazy, it is likely that similar statistics are correlated with reflectance.

6.2.2 Synthesizing material appearance

A direction for immediate future research is the problem of synthesizing material appearance. In Chapter 4 we saw that modifying the Heeger Bergen texture synthesis algorithm gave us reasonably good results. In Chapter 5 in our reflectance perception experiments we found that the synthesized material images were rated identically to the real images.

If we analyze the results in Figure 4-14 we observe that there is considerable room for improvement. The problem of synthesizing material appearance can be summarized thus - given an image of a surface of some material, a material synthesis algorithm should produce an output image of the same surface but with altered material properties. For example given an image of matte crumpled paper, we want to synthesizes images of glossy crumpled paper or wet crumpled paper. The folds of the paper and the illumination on the paper should appear to be the same.

Such an image manipulation is easy in the forward rendering framework of computer graphics where the model for surface reflectance can be tweaked. However for material synthesis, reflectance parameters of the surface have be estimated explicitly or implicitly and then manipulated.

Material synthesis can be considered a special case of the more general texture

synthesis problem. As we saw in Chapter 4, texture synthesis algorithms cannot be applied directly and need modifications. In fact using a more sophisticated texture synthesis algorithm than Heeger-Bergen, leads to poorer results. This happens because textures are stochastic and two samples of the same texture (if they are 3D) need not have the same surface structure. Therefore algorithms tailored for general texture synthesis cannot be applied directly to this problem.

Appendix A

Experiment II results

The detailed results of Experiments IIA, IIB and IIC are presented here.

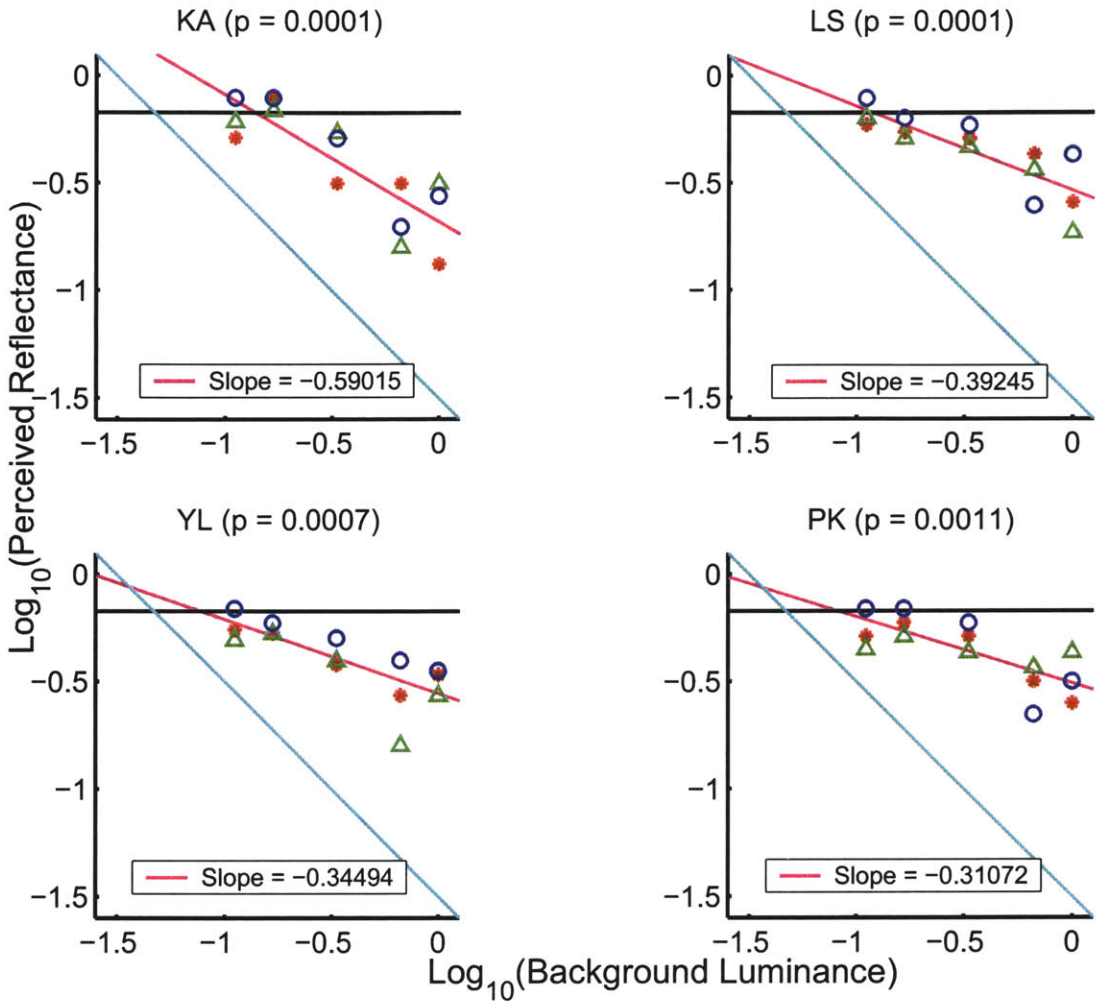


Figure A-1: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Background luminance) (Material 10, R channel)**. Luminance of material image is held fixed at 0.33 while the luminance of the background changes. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log background luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope $= -1$. The magenta line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers the slope of the fit is significantly different from 0 and -1.

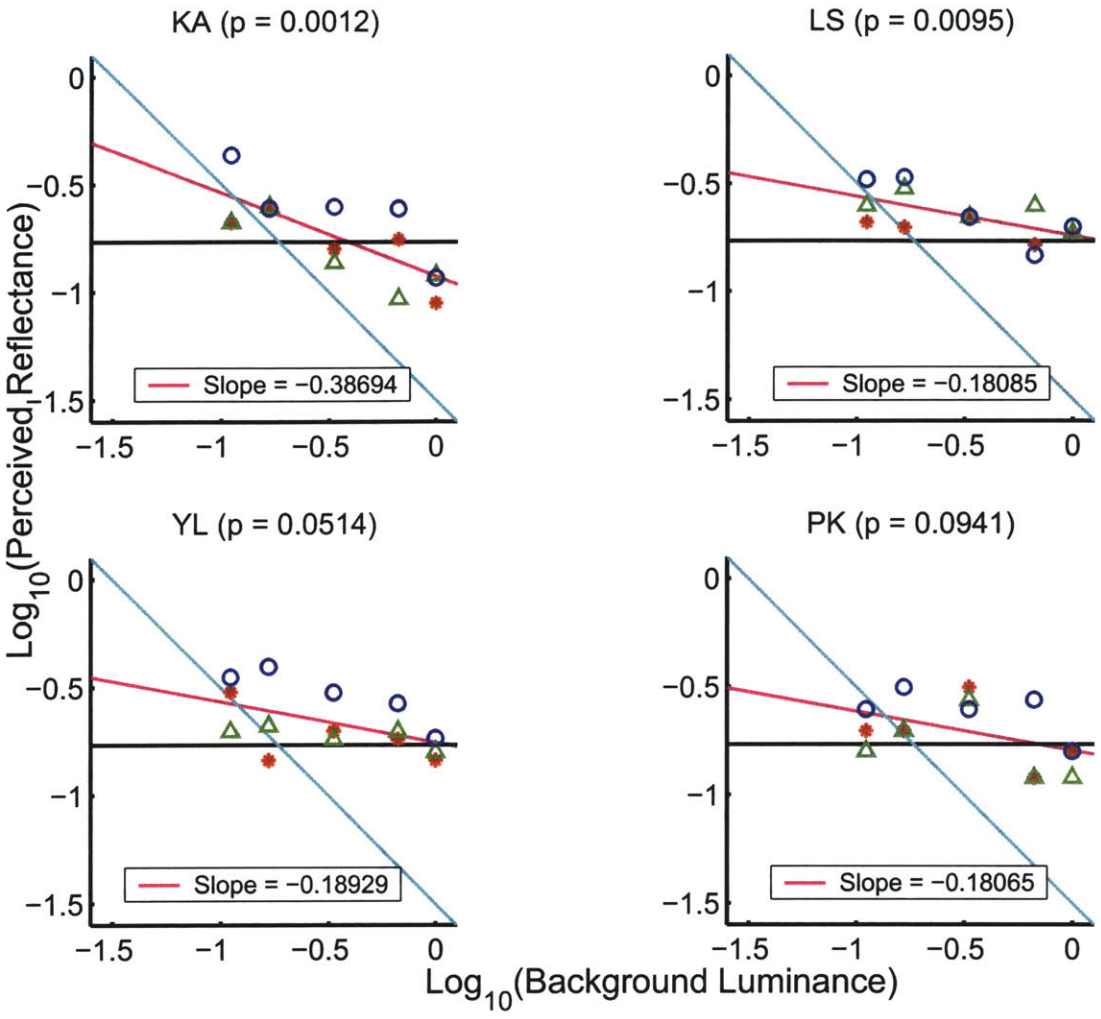


Figure A-2: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Background luminance) (Material 10, G channel)**. Luminance of material image is held fixed at 0.33 while the luminance of the background changes. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log background luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope $= -1$. The magenta line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers, except PK, the slope of the fit is significantly different from 0 and -1.

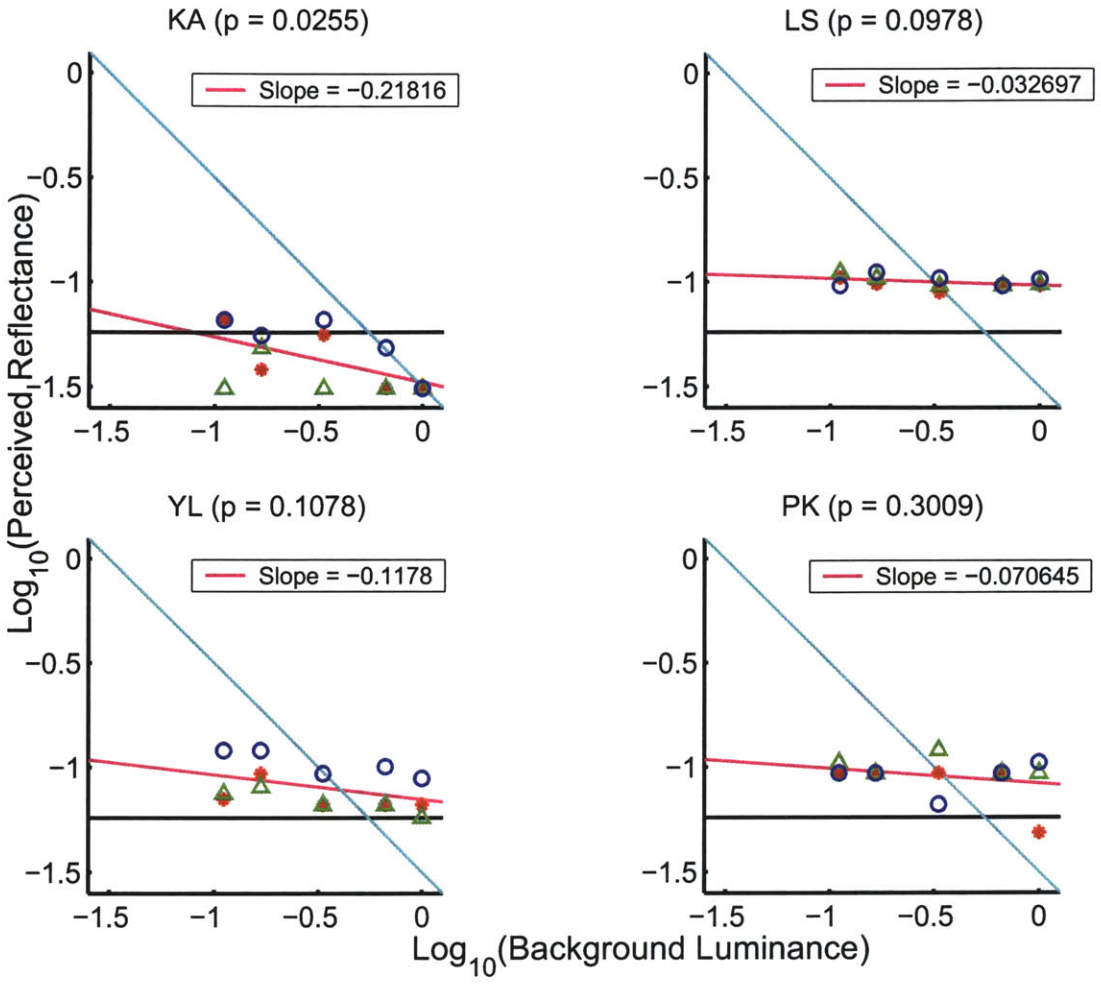


Figure A-3: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Background luminance) (Material 10, B channel)**. Luminance of material image is held fixed at 0.33 while the luminance of the background changes. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log background luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope $= -1$. The magenta line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers, except KA, the slope of the fit is not significantly different from 0.

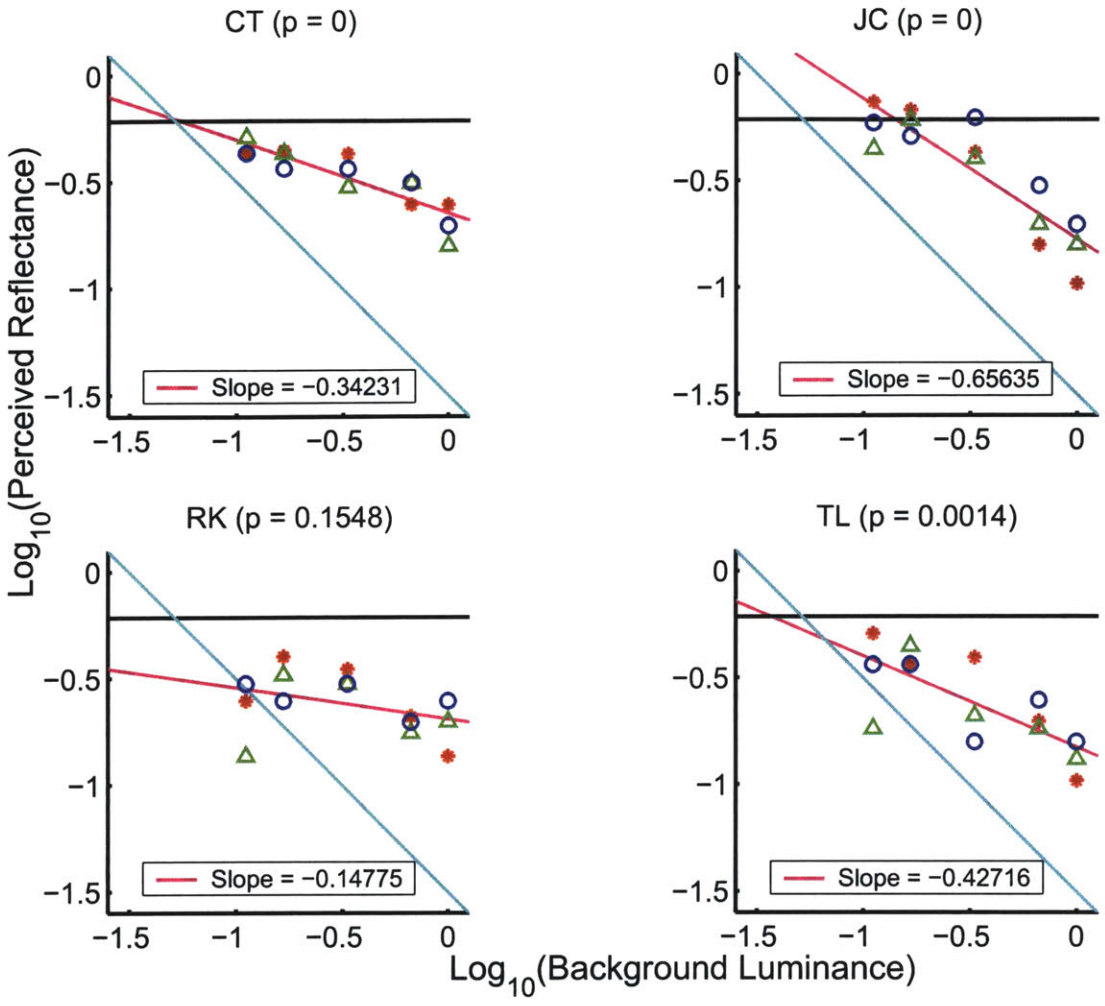


Figure A-4: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Background luminance) (Material 5, R channel)**. Luminance of material image is held fixed at 0.33 while the luminance of the background changes. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log background luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope $= -1$. The magenta line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers, except RK, the slope of the fit is significantly different from 0 and -1.

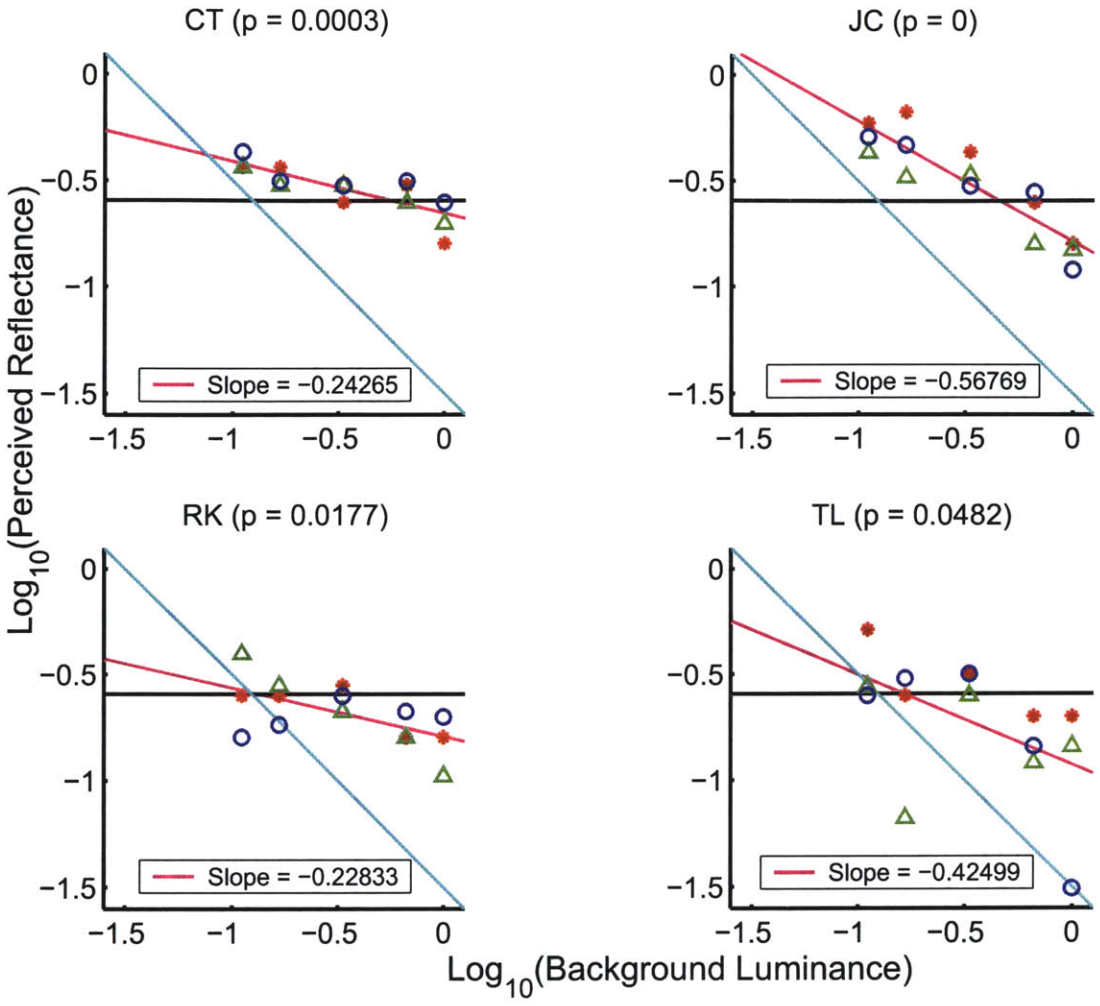


Figure A-5: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Background luminance) (Material 5, G channel)**. Luminance of material image is held fixed at 0.33 while the luminance of the background changes. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log background luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope $= -1$. The magenta line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers the slope of the fit is significantly different from 0 and -1.

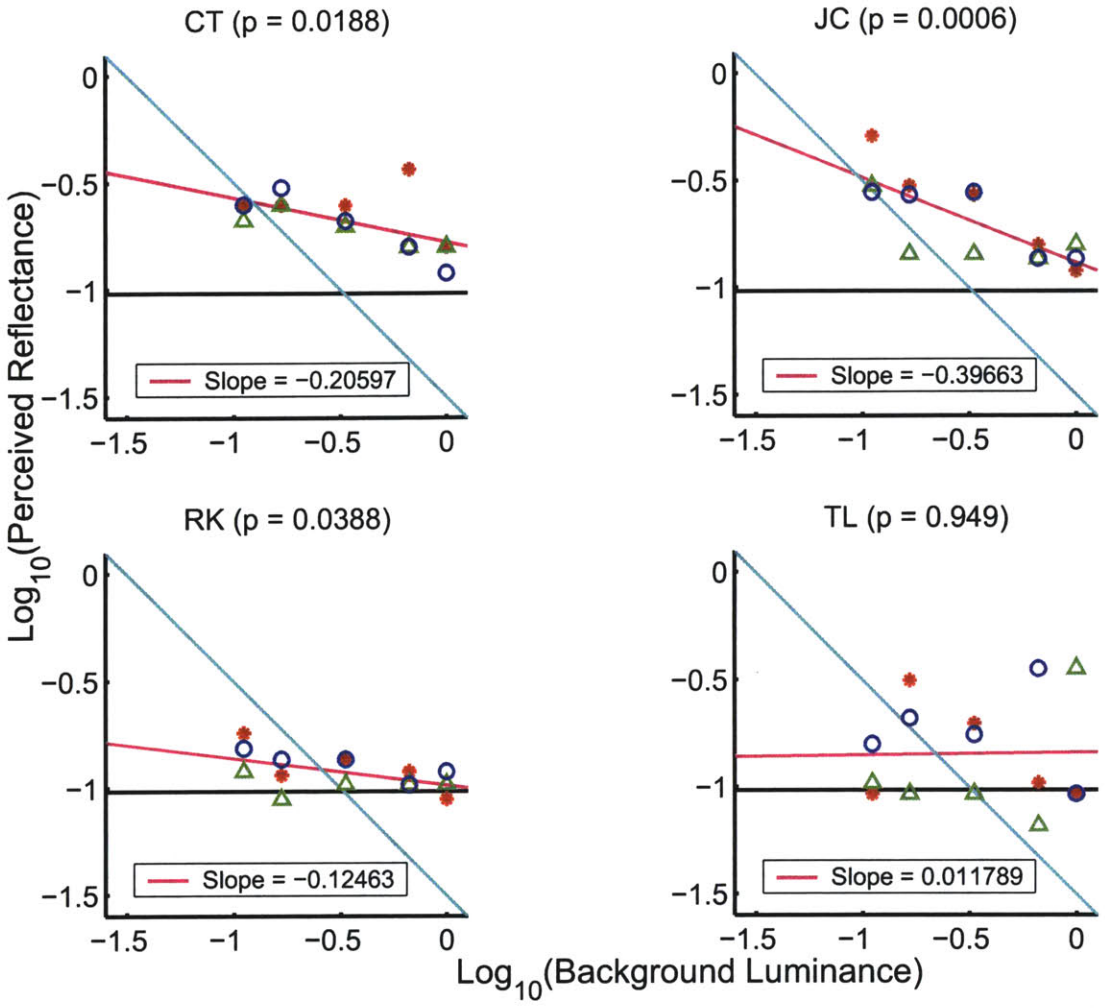


Figure A-6: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Background luminance) (Material 5, B channel)**. Luminance of material image is held fixed at 0.33 while the luminance of the background changes. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log background luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope $= -1$. The magenta line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers, except TL, the slope of the fit is significantly different from 0 and -1.

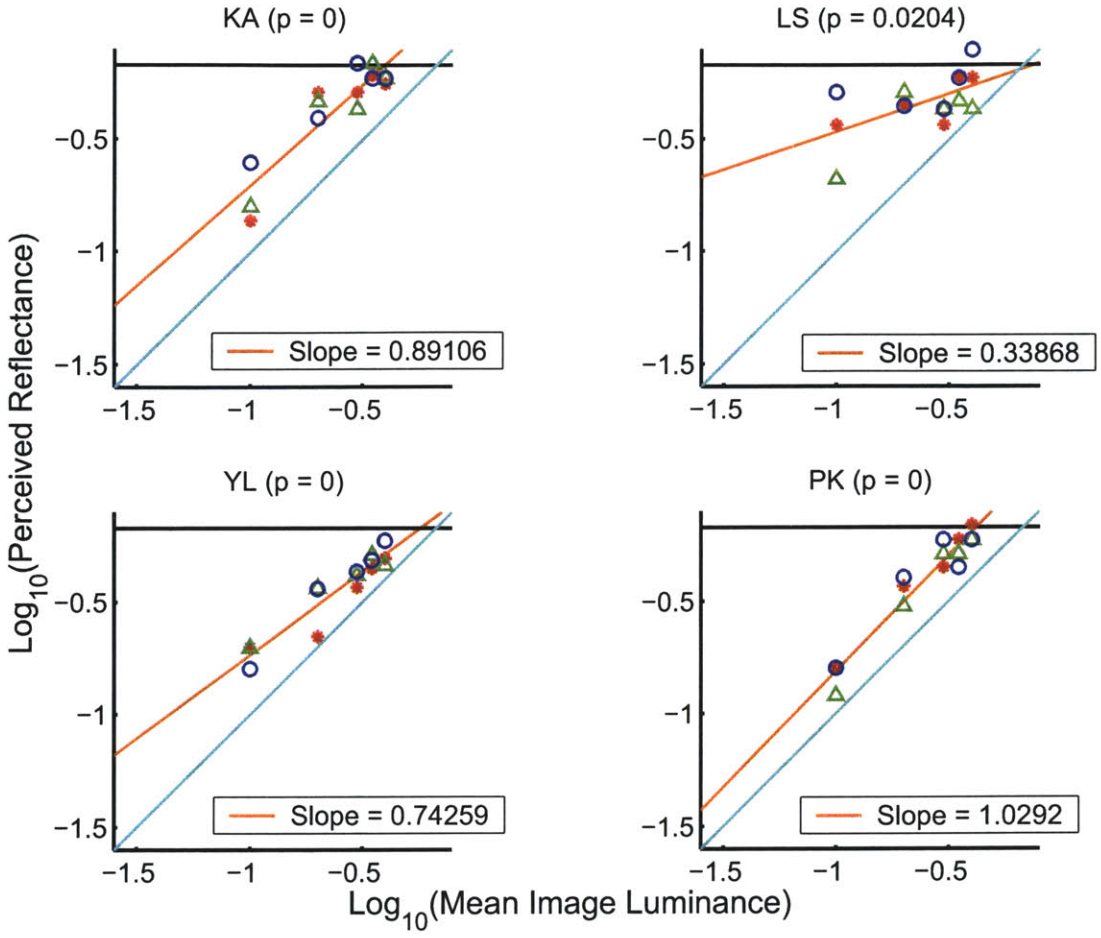


Figure A-7: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Mean Image luminance) (Material** 10**, R channel)**. Luminance of background is held fixed at 0.33 while the mean luminance of the material image changes. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log mean image luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope = 1 . The orange line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers, except PK, the slope of the fit is significantly different from 0 and 1.

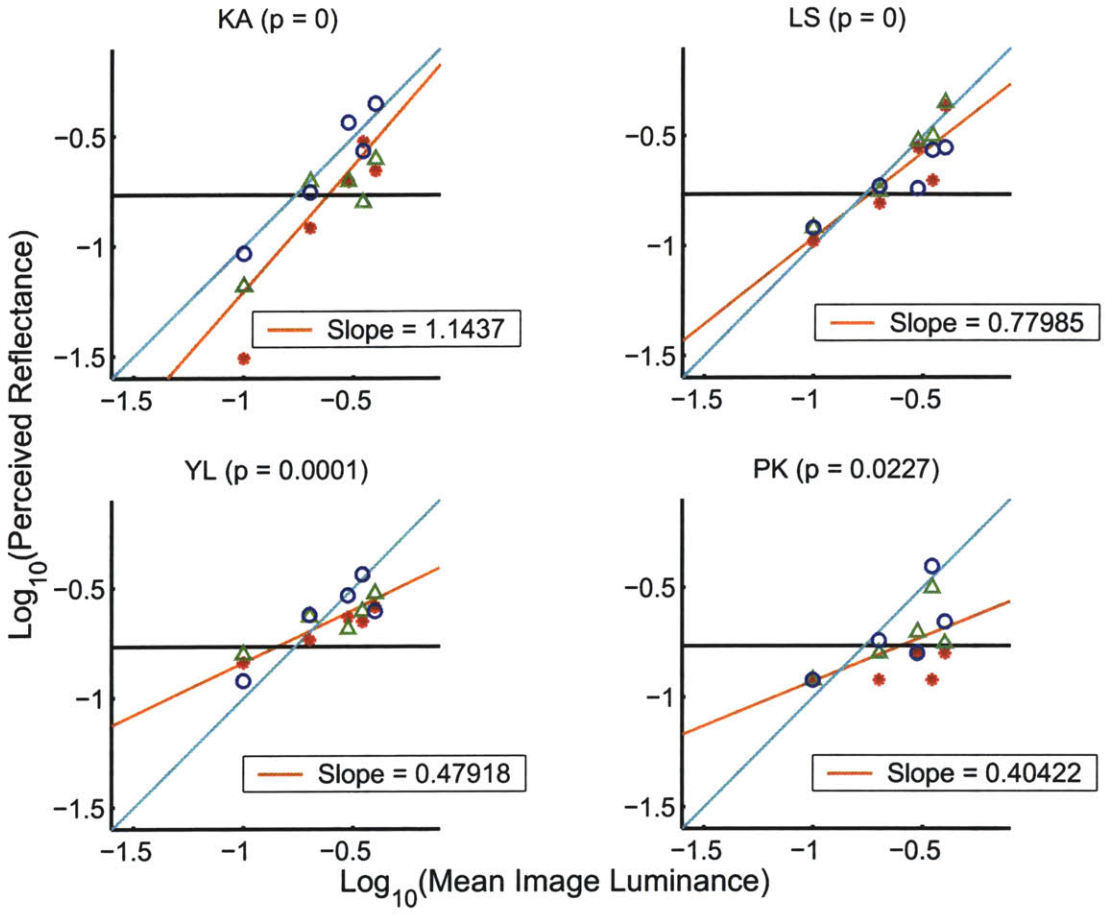


Figure A-8: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Mean Image luminance) (Material 10, G channel)**. Luminance of background is held fixed at 0.33 while the mean luminance of the material image changes. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log mean image luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope = 1 . The orange line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers, except KA, the slope of the fit is significantly different from 0 and 1.

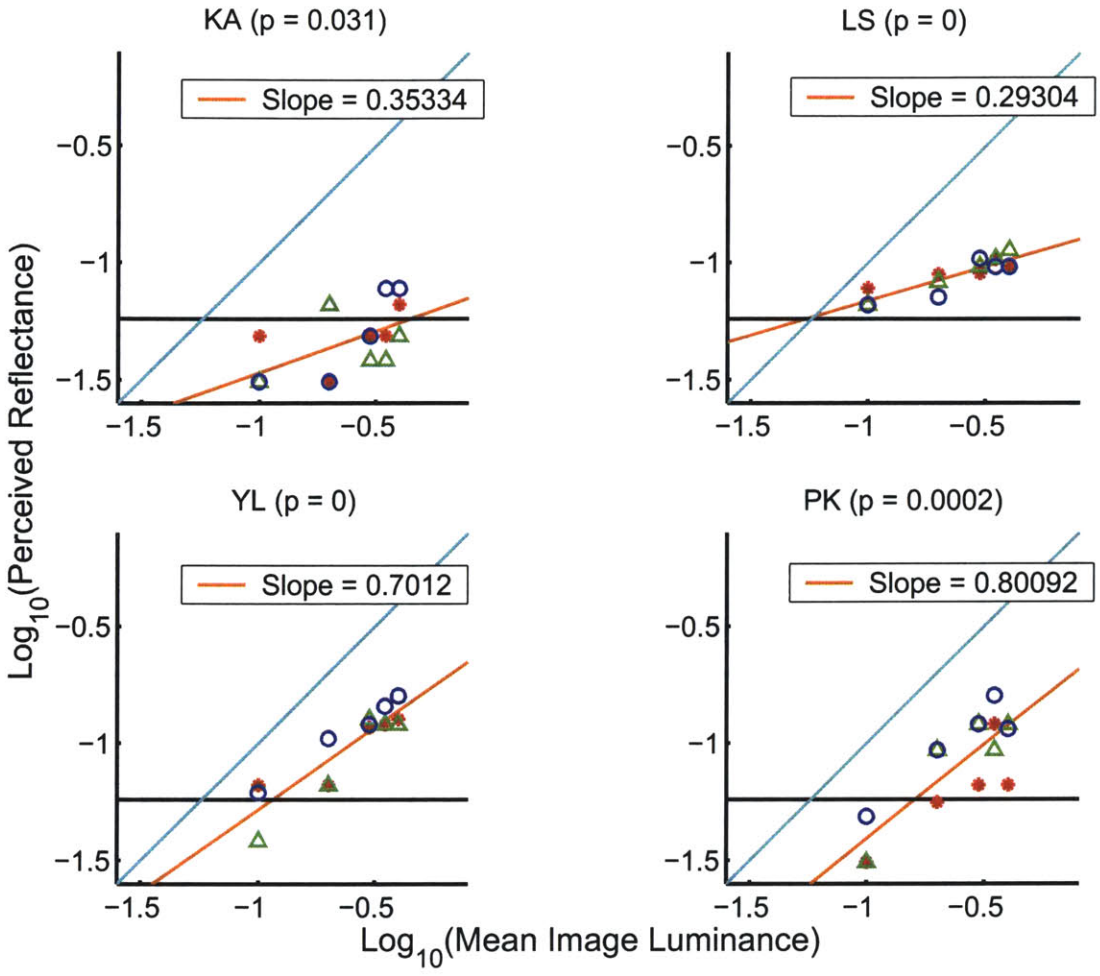


Figure A-9: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Mean Image luminance) (Material 10, B channel)**. Luminance of background is held fixed at 0.33 while the mean luminance of the material image changes. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log mean image luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope = 1 . The orange line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers the slope of the fit is significantly different from 0 and 1.

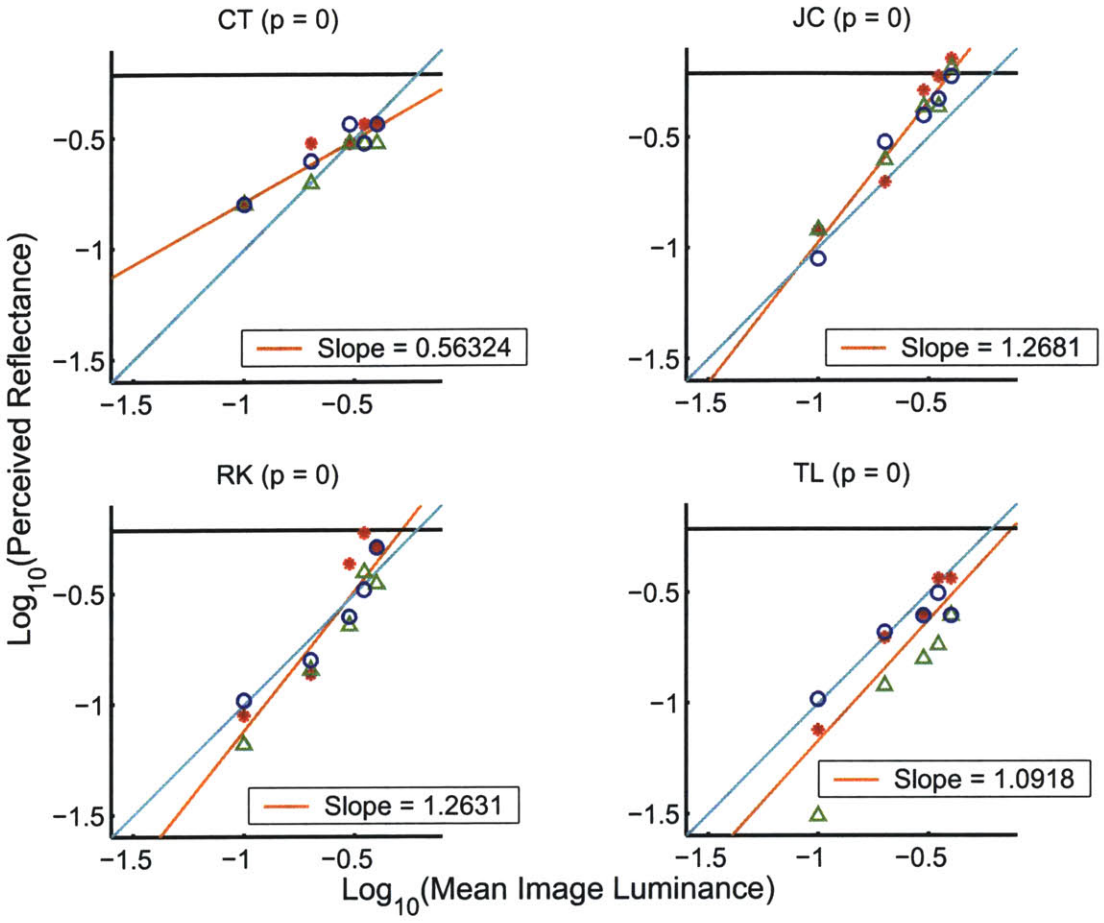


Figure A-10: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Mean Image luminance) (Material 5, R channel)**. Luminance of background is held fixed at 0.33 while the mean luminance of the material image changes. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log mean image luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope = 1 . The orange line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For two observers, RK and TL, the slope of the fit is significantly close to 1 but for the other observers, the slope is significantly different from 0 or 1.

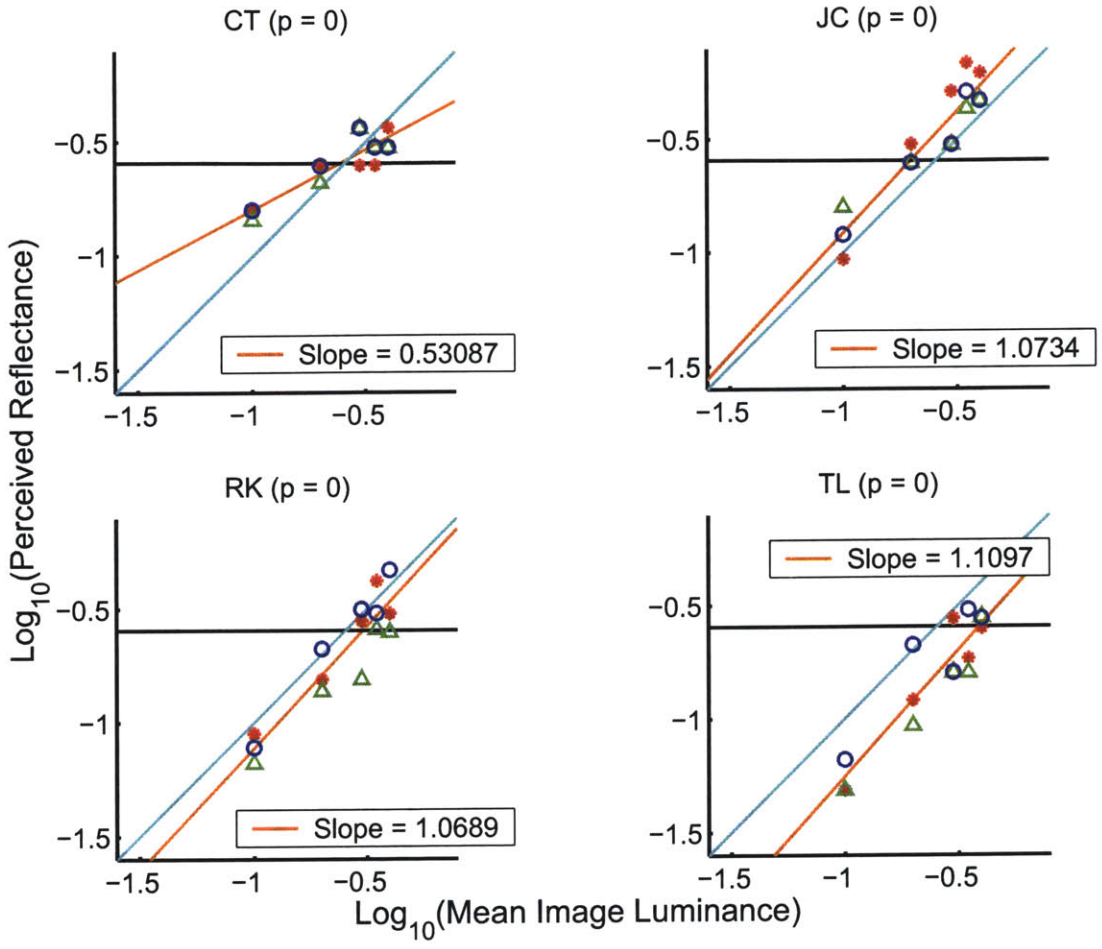


Figure A-11: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Mean Image luminance) (Material 5, G channel)**. Luminance of background is held fixed at 0.33 while the mean luminance of the material image changes. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log mean image luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope = 1 . The orange line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers, except CT, the slope of the fit is significantly close to 1.

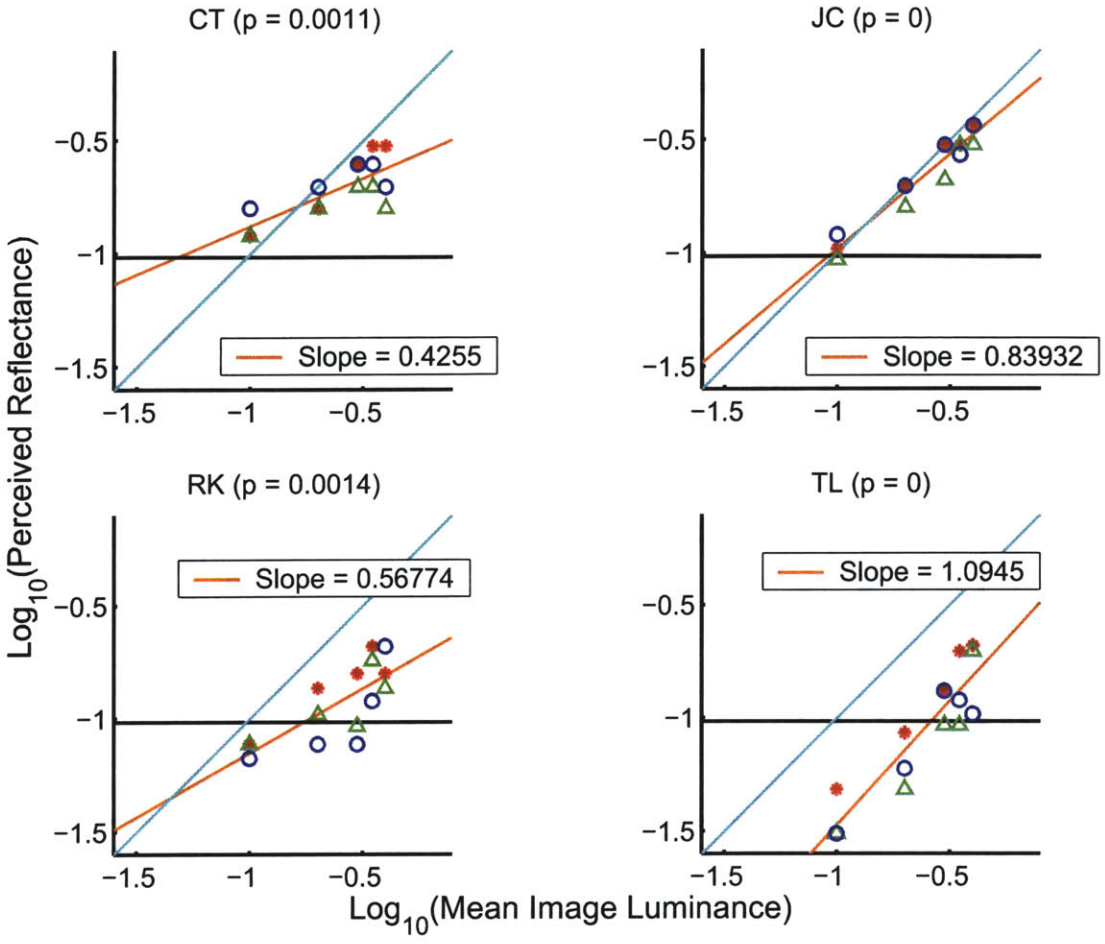


Figure A-12: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Mean Image luminance) (Material 5, B channel)**. Luminance of background is held fixed at 0.33 while the mean luminance of the material image changes. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log mean image luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope = 1 . The orange line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers, except TL, the slope of the fit is significantly different from 0 and 1.

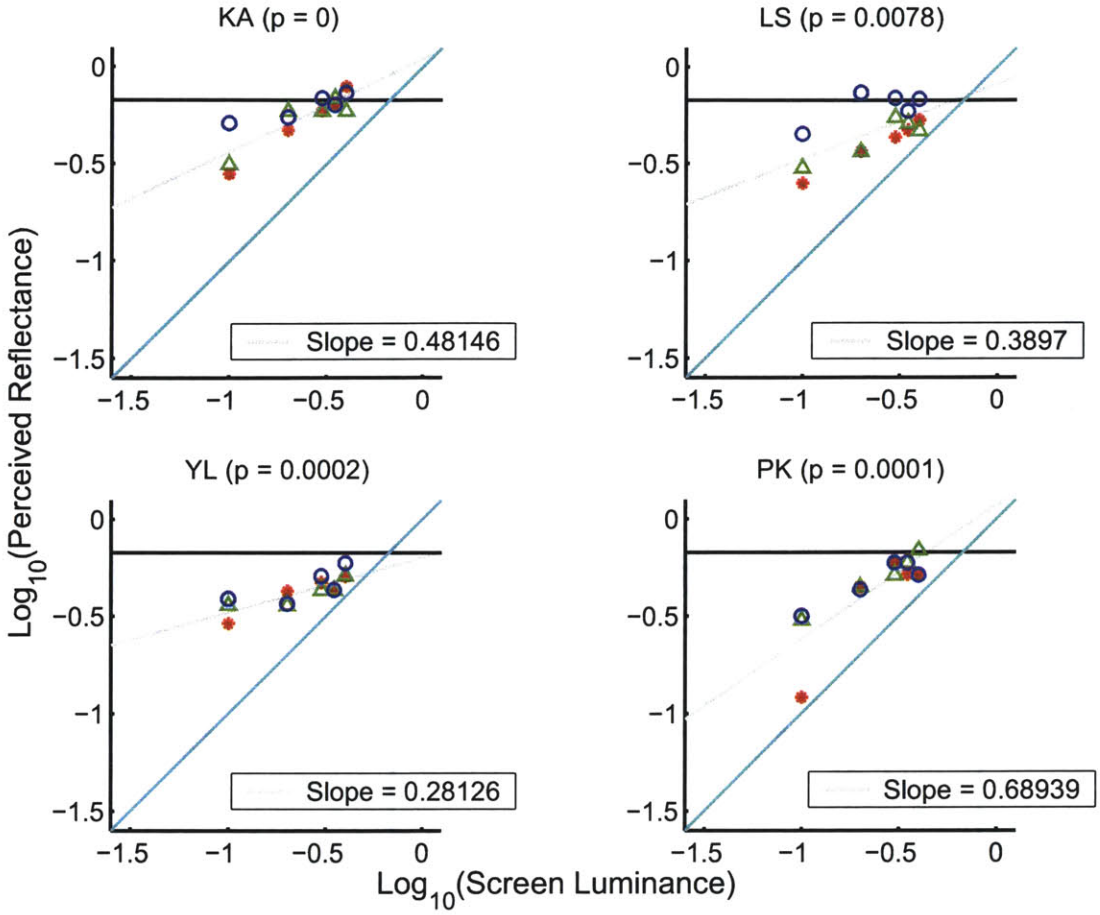


Figure A-13: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Screen luminance) (Material 10, R channel).** The ratio of the luminance of background to that of the image is held fixed at 1 while the mean luminances of the material image and background change. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log screen luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope = 1 . The gray line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers the slope of the fit is significantly different from 0 and 1.

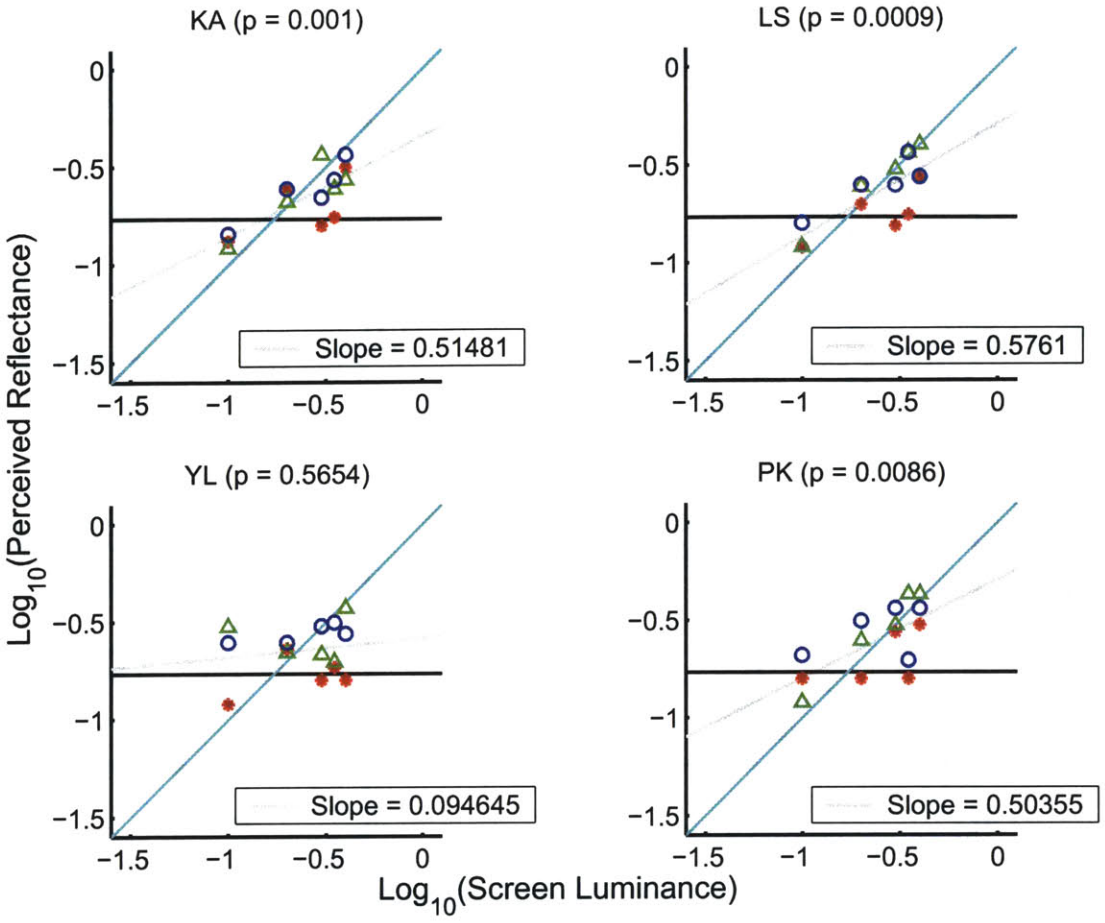


Figure A-14: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Screen luminance) (Material 10, G channel).** The ratio of the luminance of background to that of the image is held fixed at 1 while the mean luminances of the material image and background change. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log screen luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope = 1 . The gray line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers, except YL, the slope of the fit is significantly different from 0 and 1.

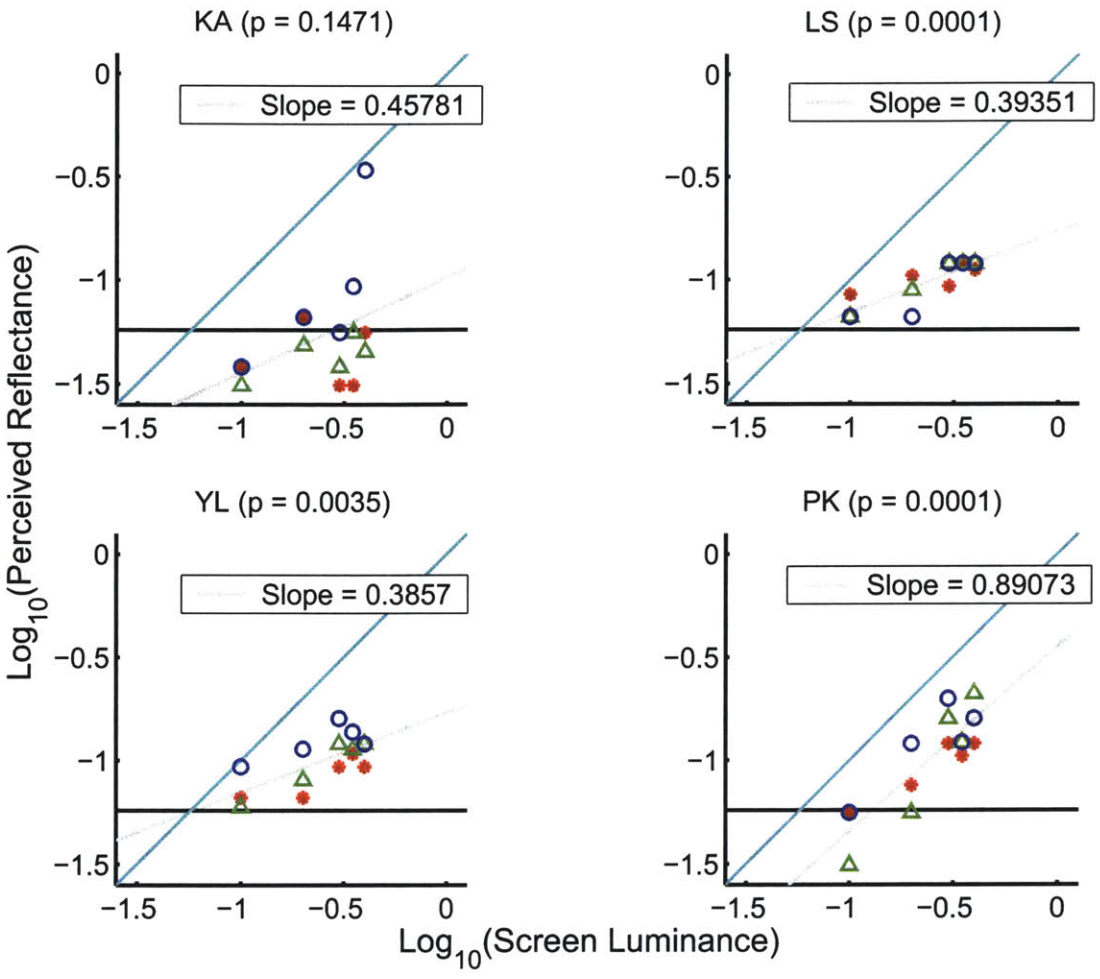


Figure A-15: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Screen luminance) (Material 10, B channel)**. The ratio of the luminance of background to that of the image is held fixed at 1 while the mean luminances of the material image and background change. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log screen luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope = 1 . The gray line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers, except KA, the slope of the fit is significantly different from 0 and 1.

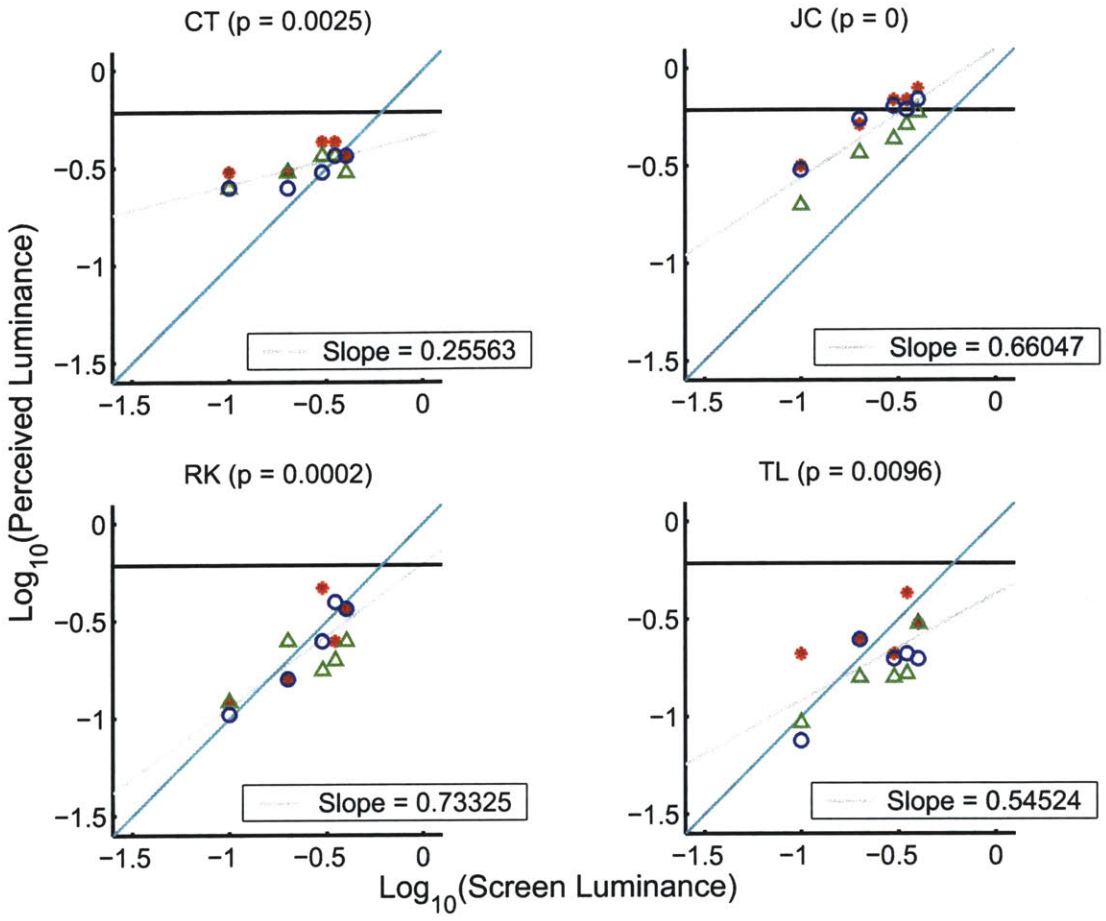


Figure A-16: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Screen luminance) (Material 5, R channel).** The ratio of the luminance of background to that of the image is held fixed at 1 while the mean luminances of the material image and background change. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log screen luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope = 1 . The gray line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers the slope of the fit is significantly different from 0 and 1.

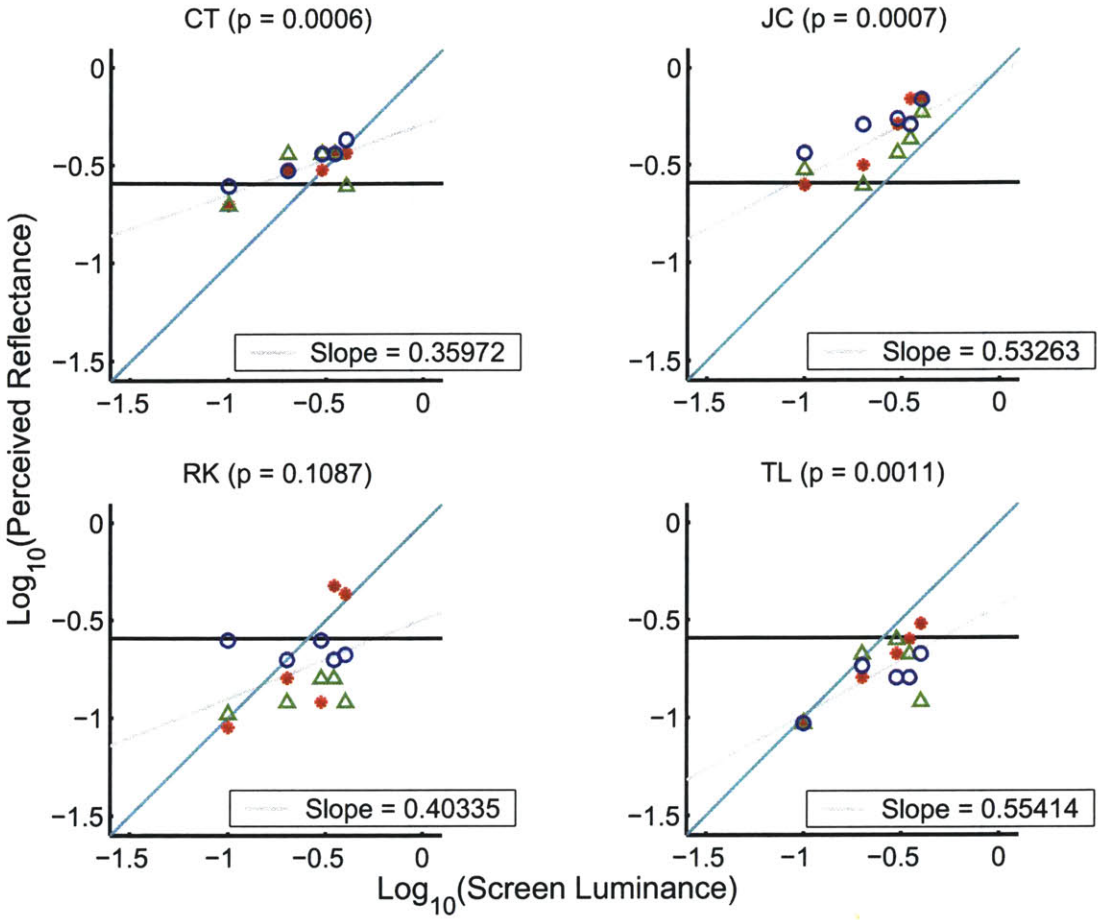


Figure A-17: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Screen luminance) (Material 5, G channel).** The ratio of the luminance of background to that of the image is held fixed at 1 while the mean luminances of the material image and background change. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log screen luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope = 1 . The gray line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers, except RK, the slope of the fit is significantly different from 0 and 1.

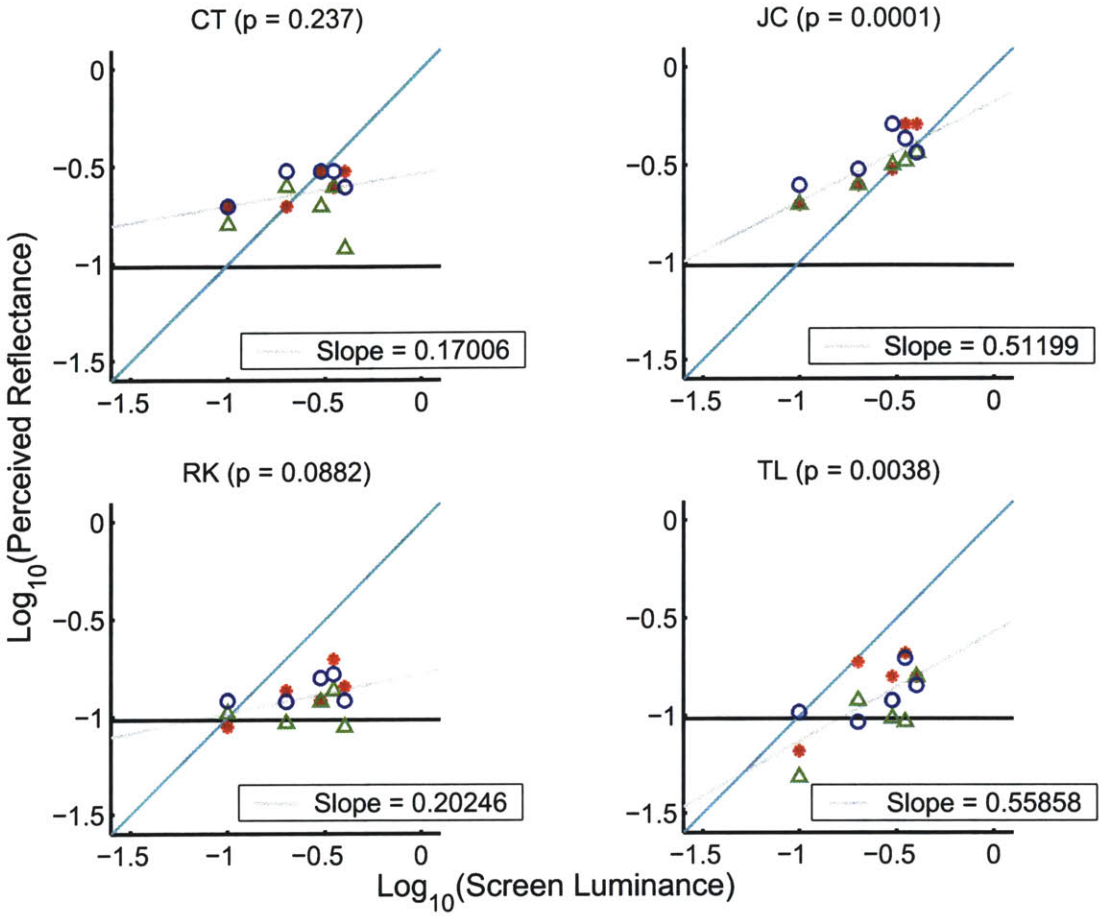


Figure A-18: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Screen luminance) (Material 5, B channel)**. The ratio of the luminance of background to that of the image is held fixed at 1 while the mean luminances of the material image and background change. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log screen luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope = 1 . The gray line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For two observers, JC and TL, the slope of the fit is significantly different from 0 and 1. For the other observers, the slope is not significant.

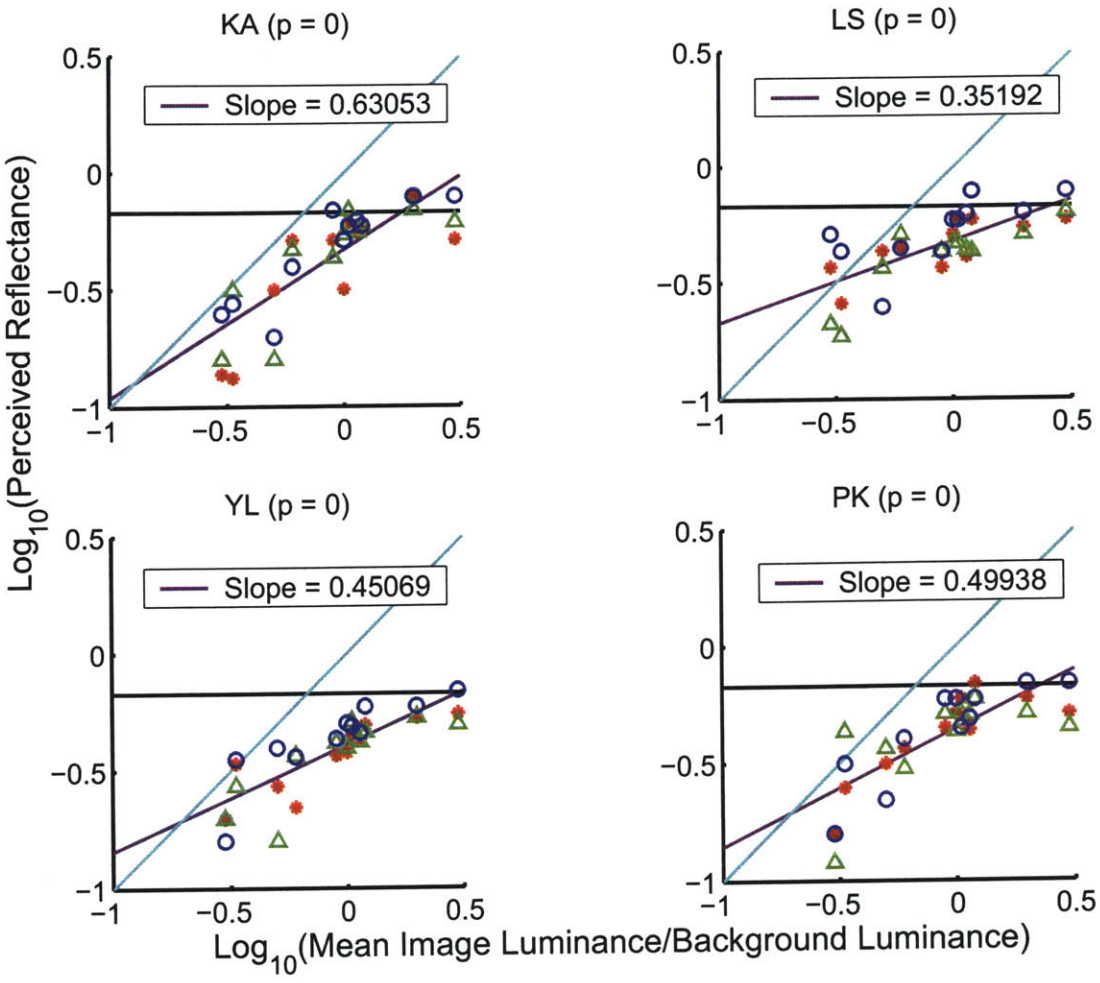


Figure A-19: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Luminance Ratio) (Material 10, R channel)**. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log of the ratio of image luminance to background luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope = 1 . The purple line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers, the slope of the fit is significantly different from 0 and 1.

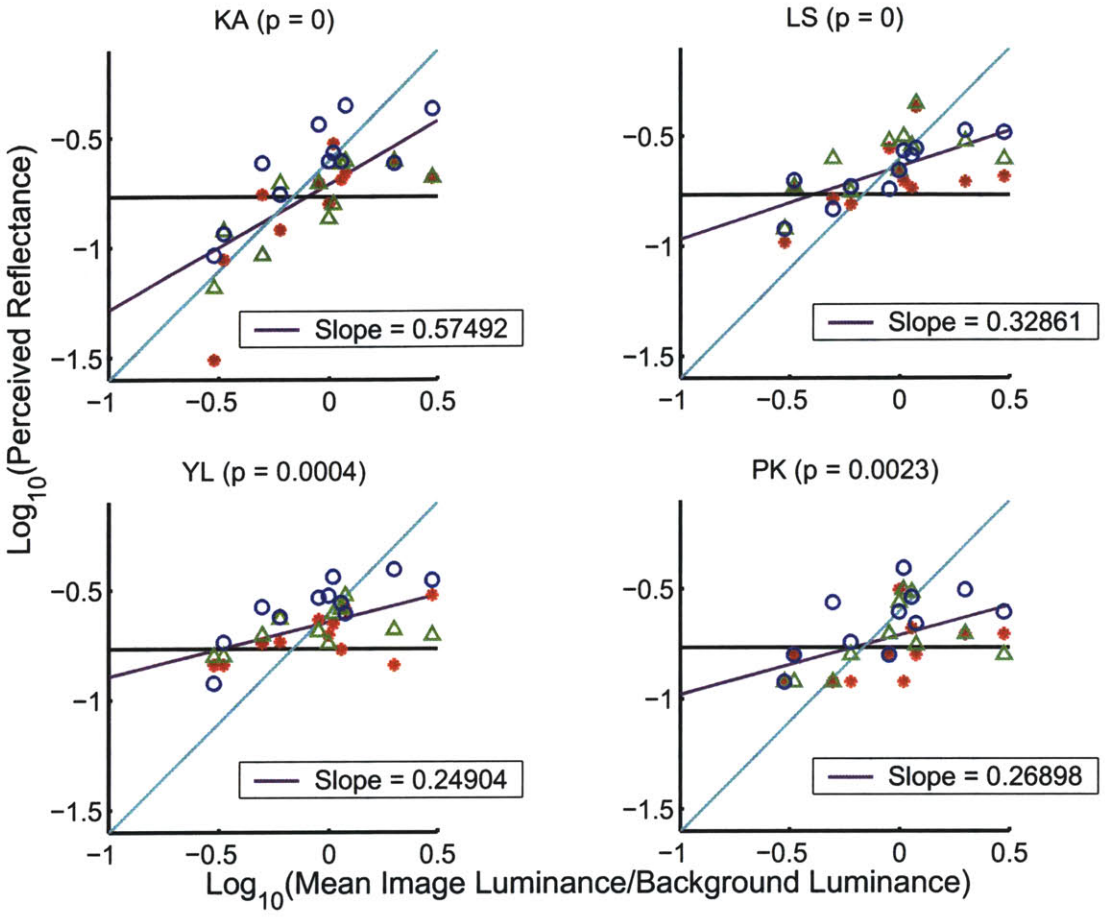


Figure A-20: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Luminance Ratio) (Material 10, G channel).** The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log of the ratio of image luminance to background luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope = 1 . The purple line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers, the slope of the fit is significantly different from 0 and 1.

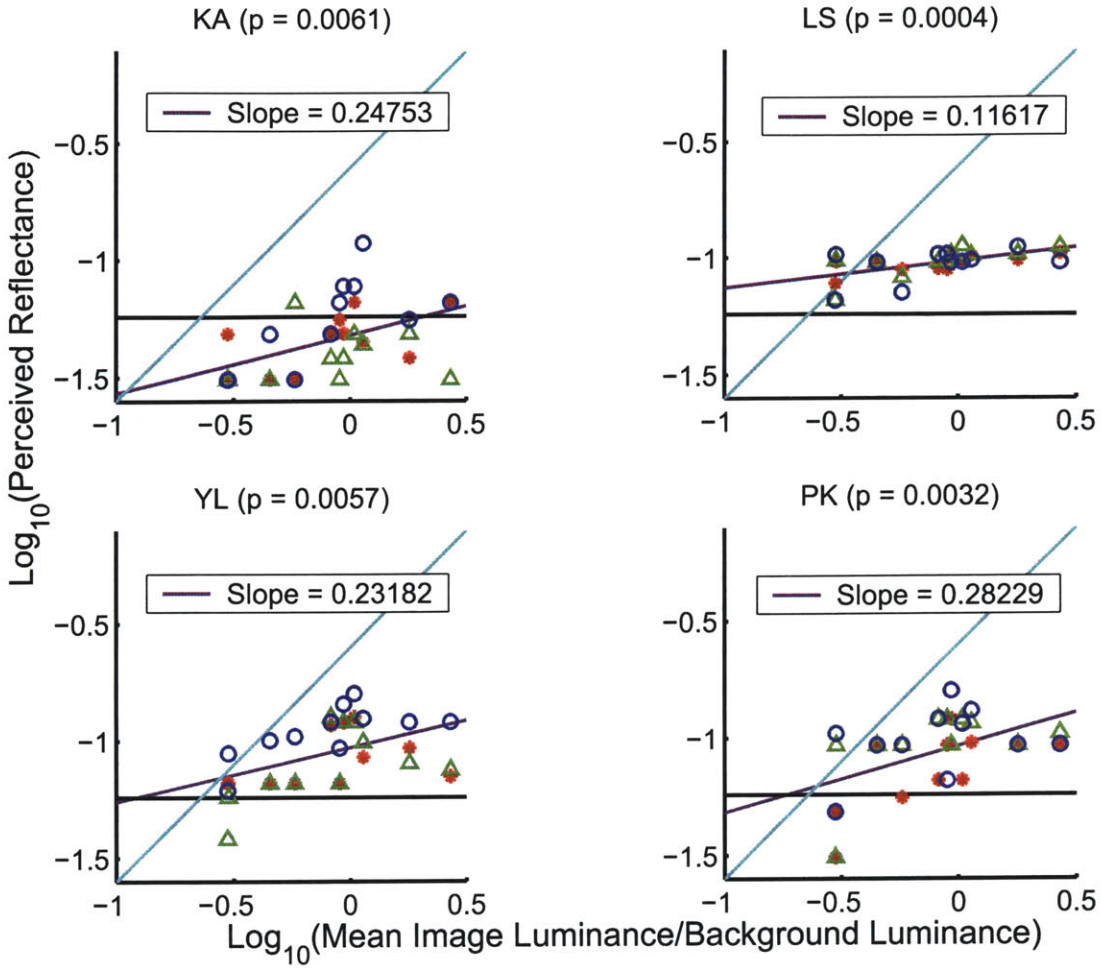


Figure A-21: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Luminance Ratio) (Material 10, B channel)**. The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log of the ratio of image luminance to background luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope = 1 . The purple line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers, the slope of the fit is significantly different from 0 and 1.

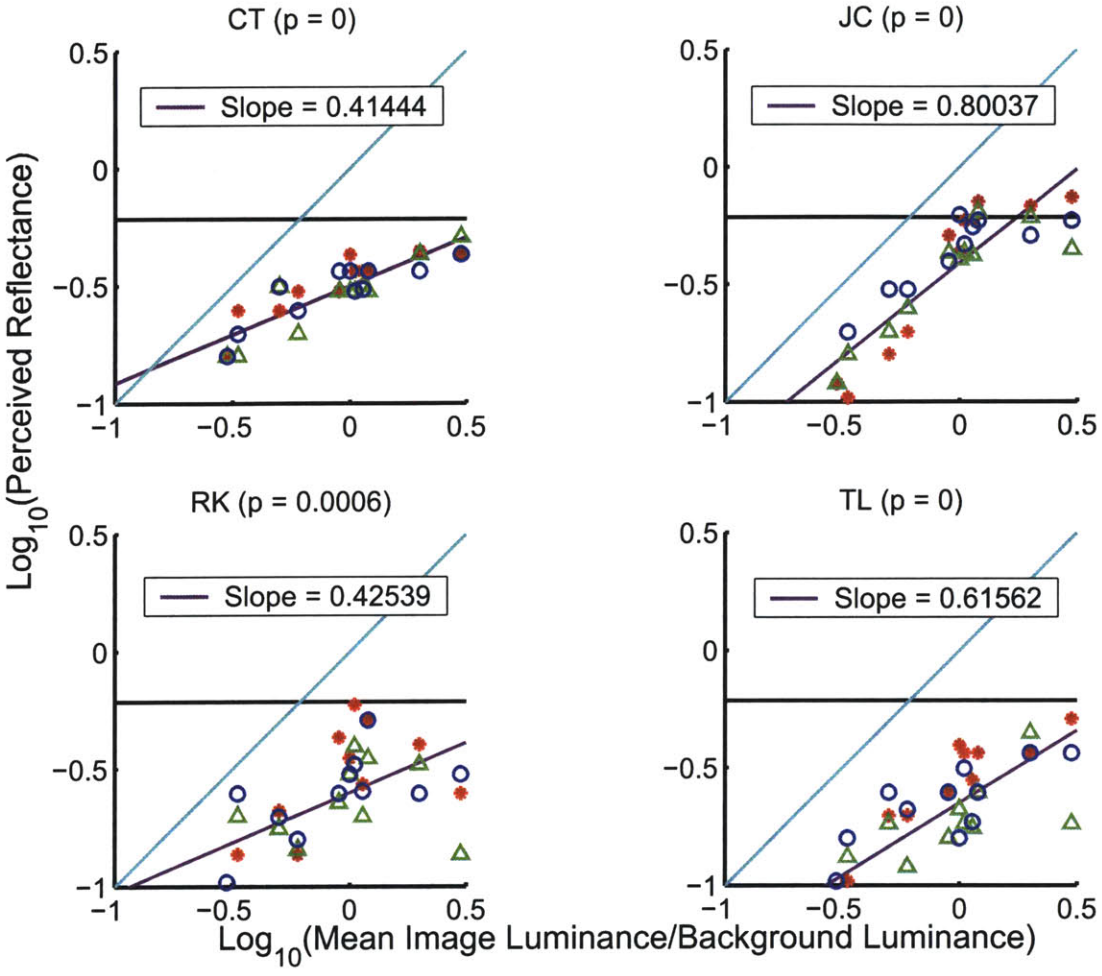


Figure A-22: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Luminance Ratio) (Material 5, R channel).** The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log of the ratio of image luminance to background luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope = 1 . The purple line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers, the slope of the fit is significantly different from 0 and 1.

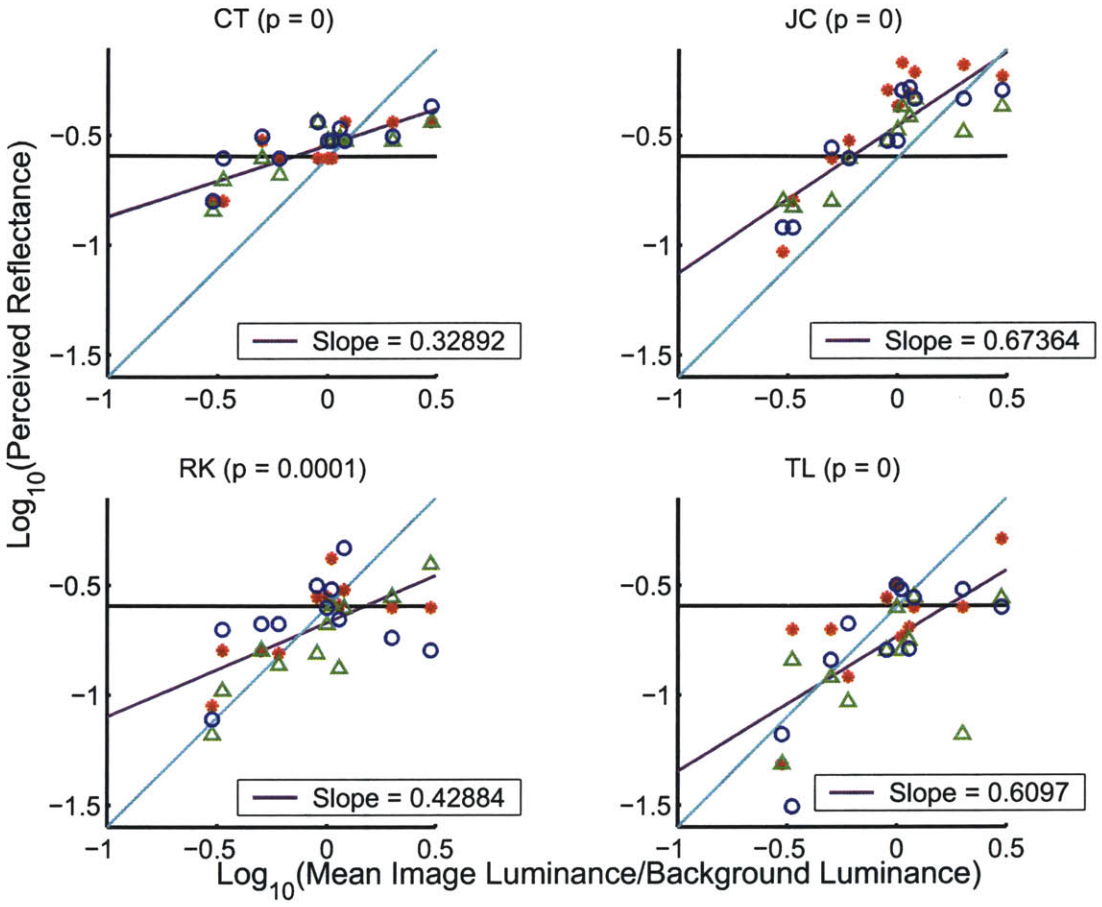


Figure A-23: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Luminance Ratio) (Material 5, G channel).** The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log of the ratio of image luminance to background luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope = 1 . The purple line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers, the slope of the fit is significantly different from 0 and 1.

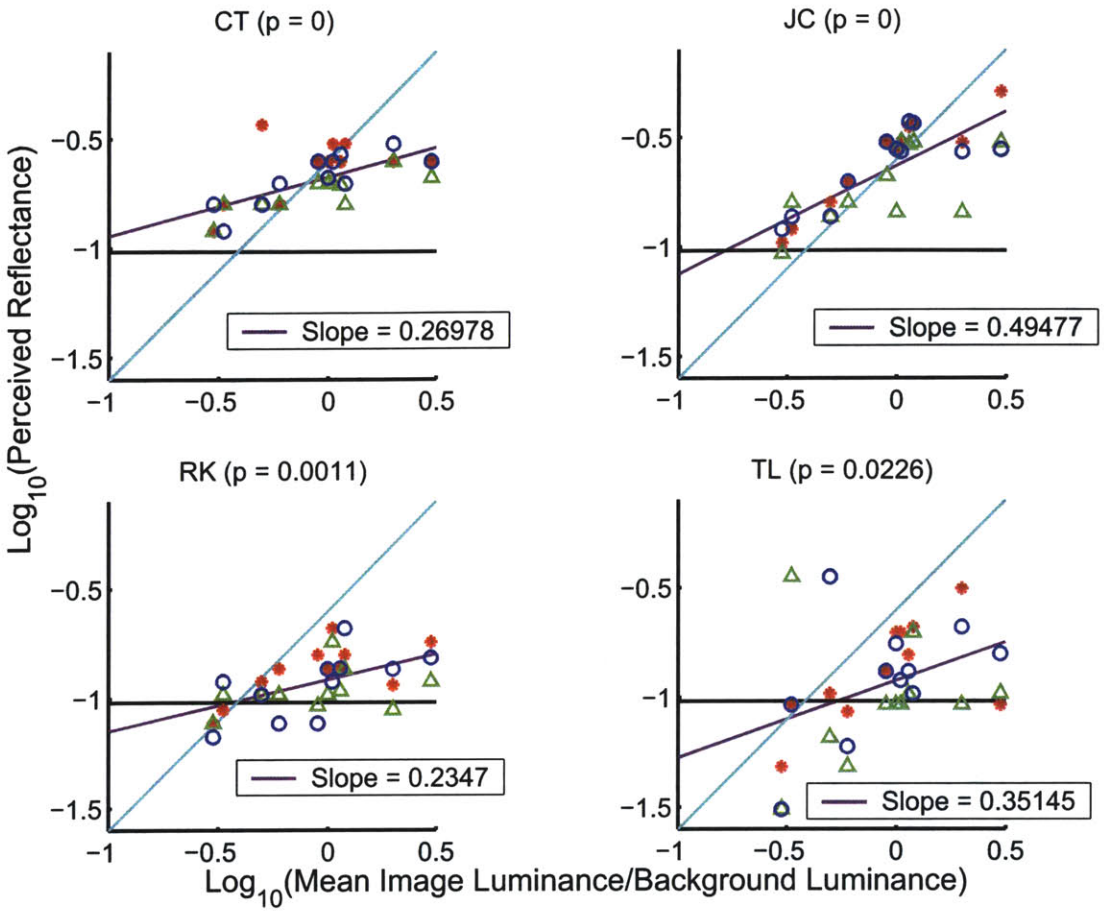


Figure A-24: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Luminance Ratio) (Material 5, B channel).** The mean log responses for each light condition (Red = Light 1, Green = Light 2, Blue = Light 3) are plotted against the log of the ratio of image luminance to background luminance for each observer. The responses of a veridical observer would lie along the horizontal ground truth line (black). If an observer demonstrates zero constancy and follows the ratio hypothesis, the responses would lie along a line parallel to the cyan line with slope = 1 . The purple line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers, the slope of the fit is significantly different from 0 and 1.

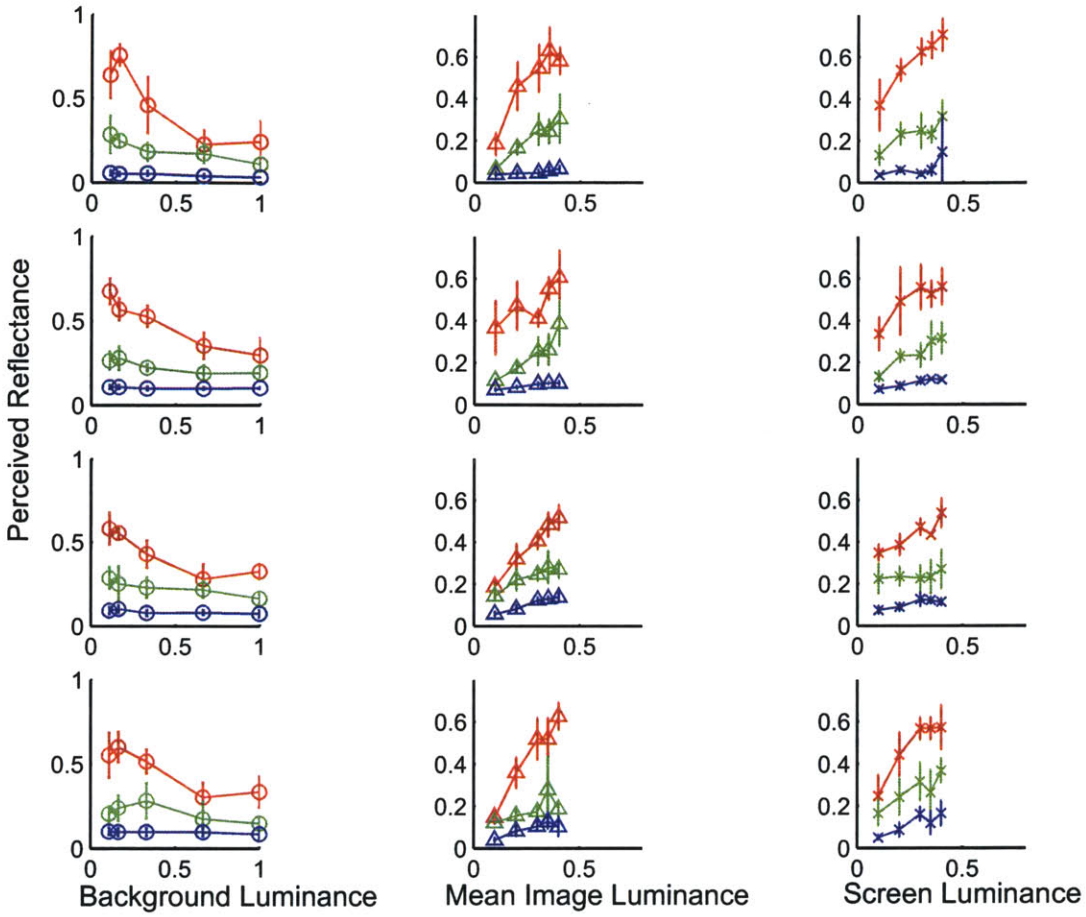


Figure A-25: **Material 10** Perceived reflectance is plotted against background luminance (keeping mean image luminance constant) in column 1, against mean luminance (keeping background luminance constant) in column 2 and against the screen luminance (keeping the ratio of image to background constant) in column 3. Each row corresponds to each of the 4 observers (KA, LS, PK and YL) who participated in these tasks. In each plot, the R channel is denoted by red, the G channel by green and B channel by blue. For each channel, the data is pooled across the 3 lighting conditions. The errorbars are the 95% confidence intervals. While the observers do not display perfect lightness constancy, they can nevertheless differentiate between the R, G and B channels for identical experimental conditions. They consistently rate the channels in the order $R > G > B$ or as in the plot $red > green > blue$.

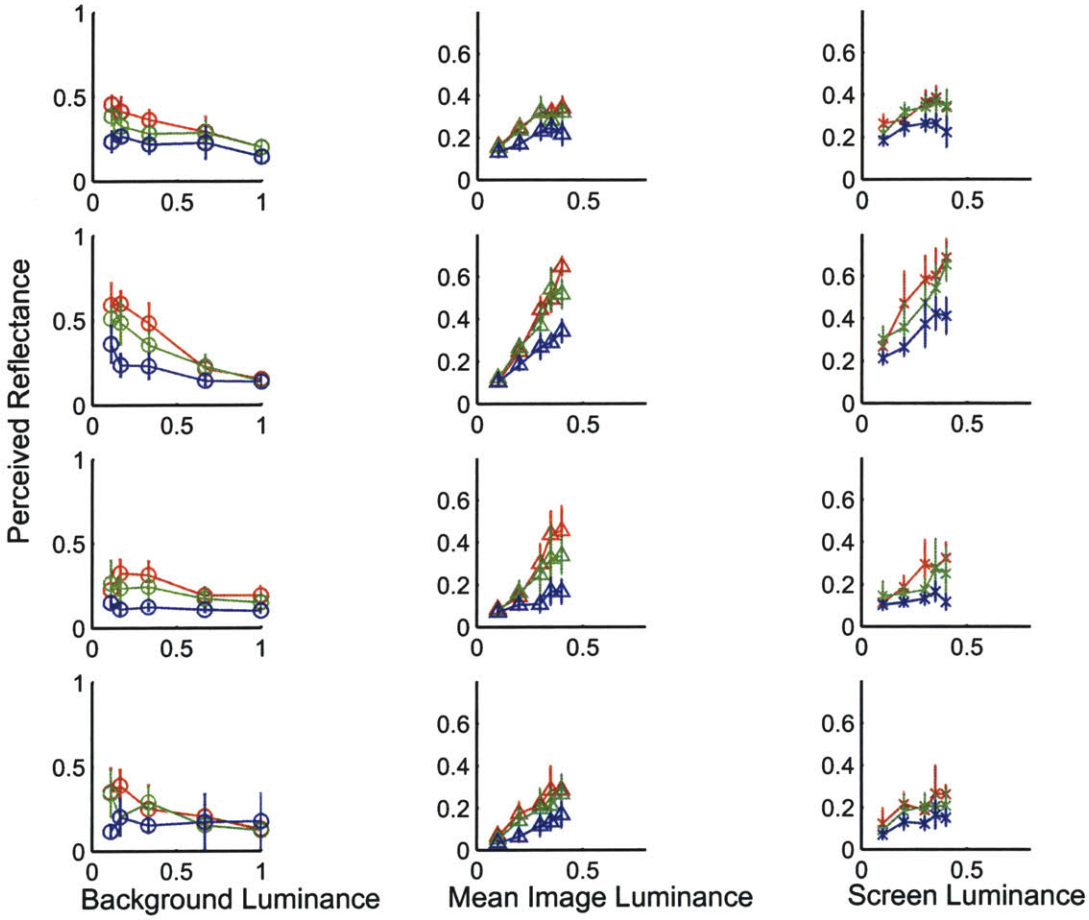
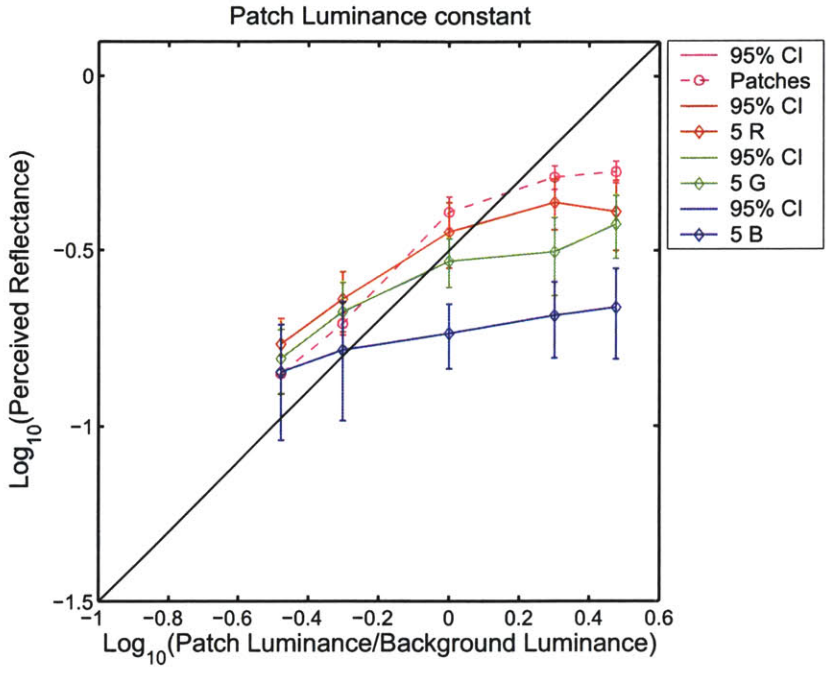
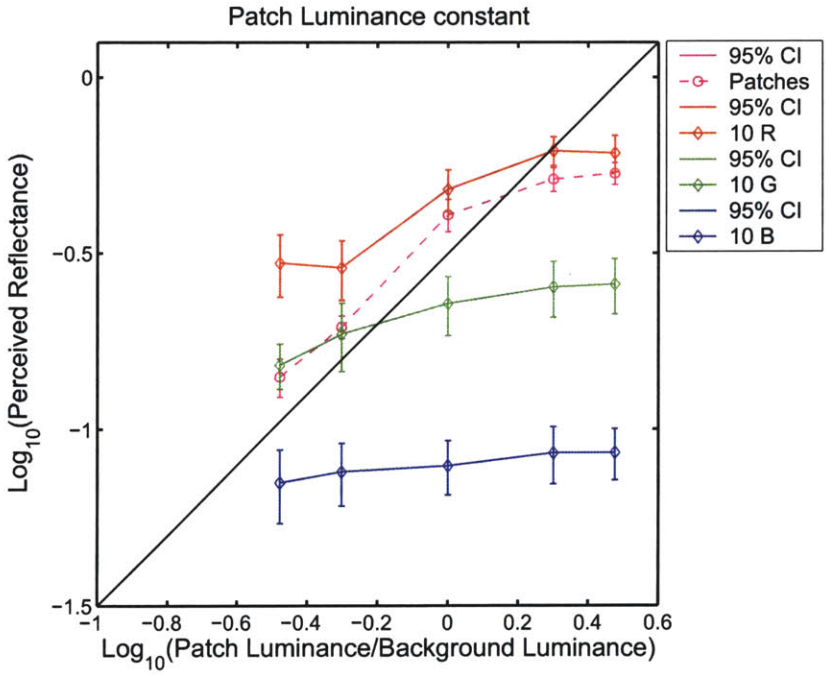


Figure A-26: **Material 5** Similar plot to Figure 5-13, except that observers viewed Material 5 instead of Material 10.

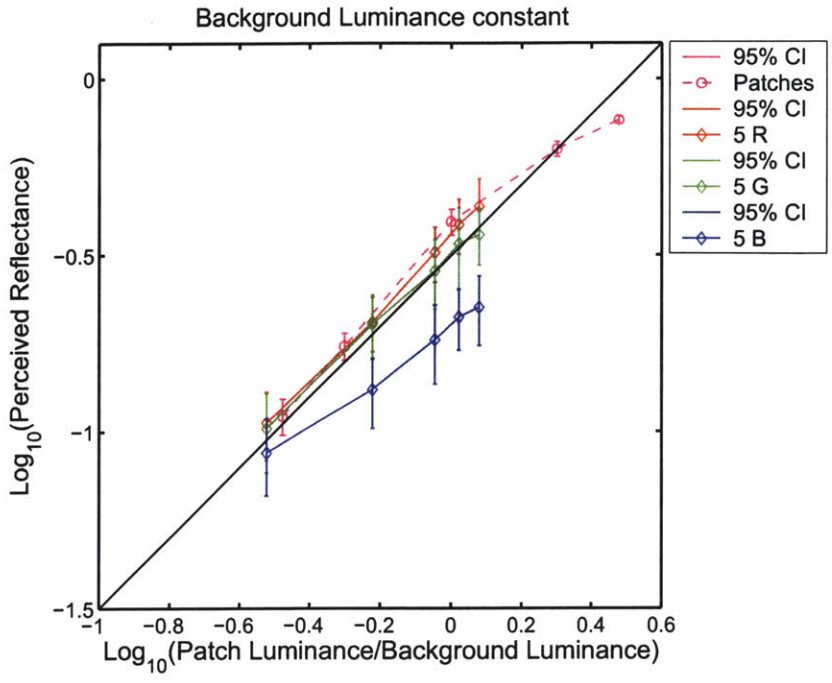


(a)

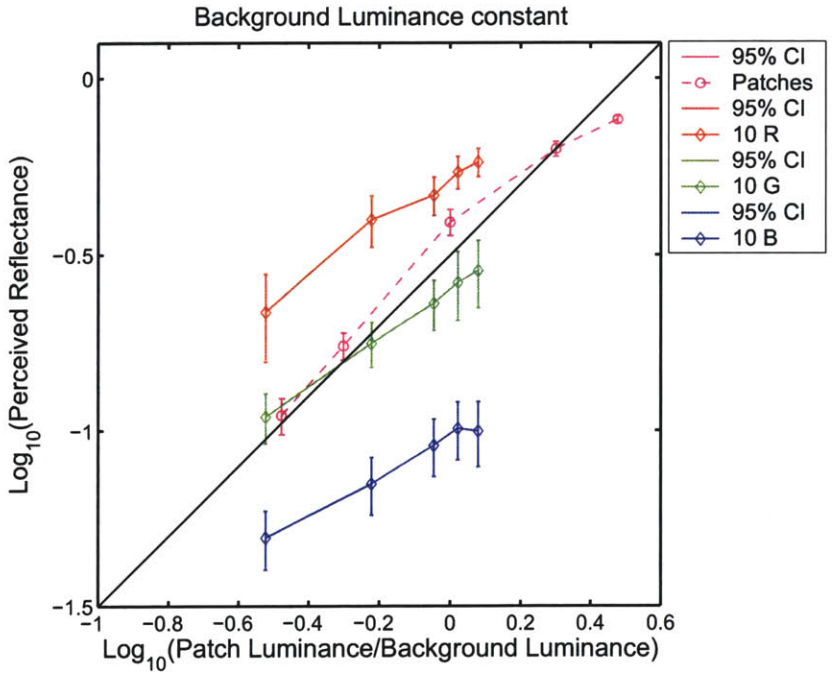


(b)

Figure A-27: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Patch Luminance/Background Luminance), patch luminance is constant** (a) Comparison with results from Experiment IIA for Material 5 and (b) for Material 10. If observers follow the ratio rule all observations would lie on a line parallel to the black line. The data is pooled across all observers and all lights.

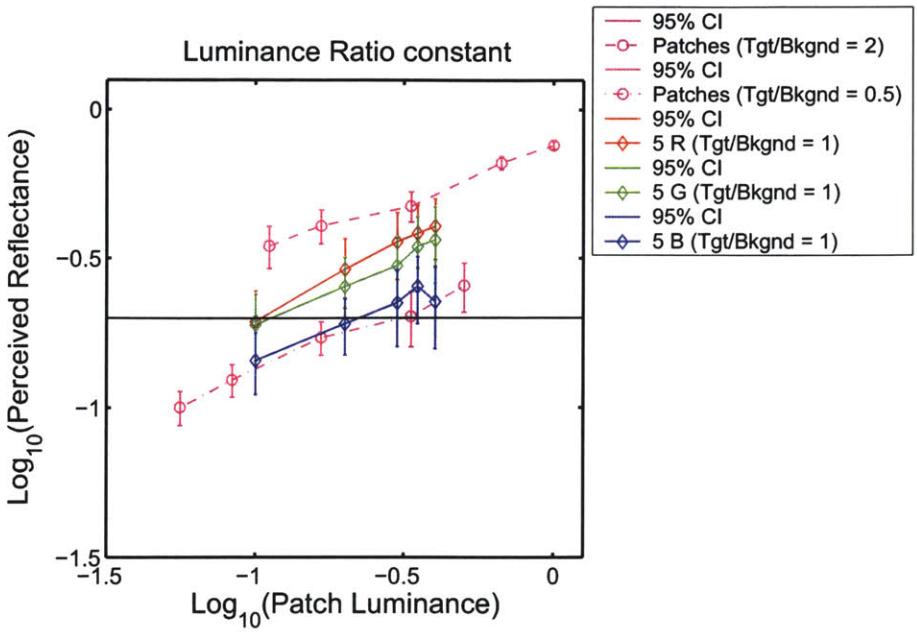


(a)

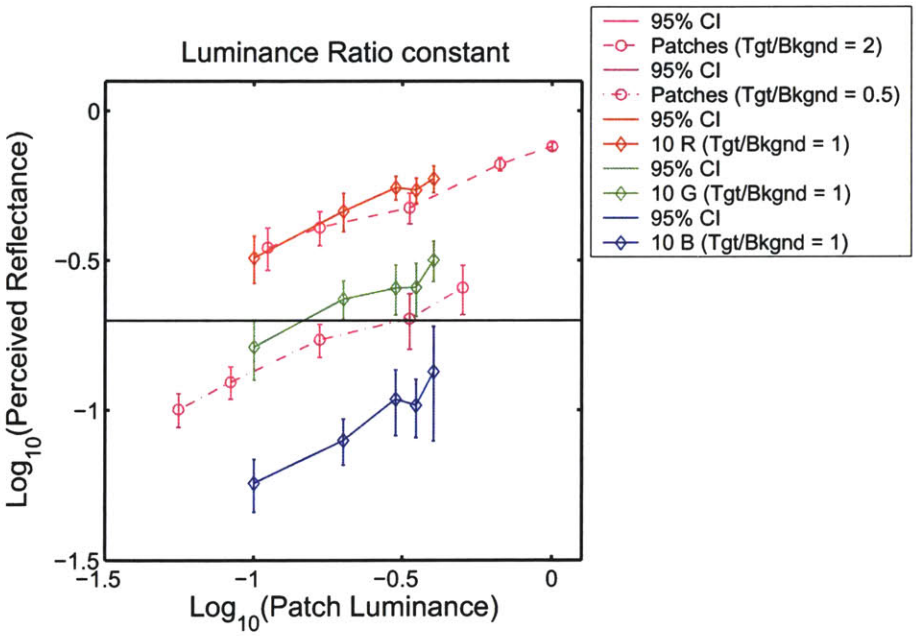


(b)

Figure A-28: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Patch Luminance/Background Luminance), background luminance is constant** (a) Comparison with results from Experiment IIA for Material 5 and (b) for Material 10. If observers follow the ratio rule all observations would lie on a line parallel to the black line. The data is pooled across all observers and all lights.



(a)



(b)

Figure A-29: Log_{10}**(Perceived reflectance) vs** Log_{10}**(Patch Luminance), luminance ratio is constant** (a) Comparison with results from Experiment IIA for Material 5 and (b) for Material 10. If observers follow the ratio rule all observations would lie on a line parallel to the black line. The data is pooled across all observers and all lights.

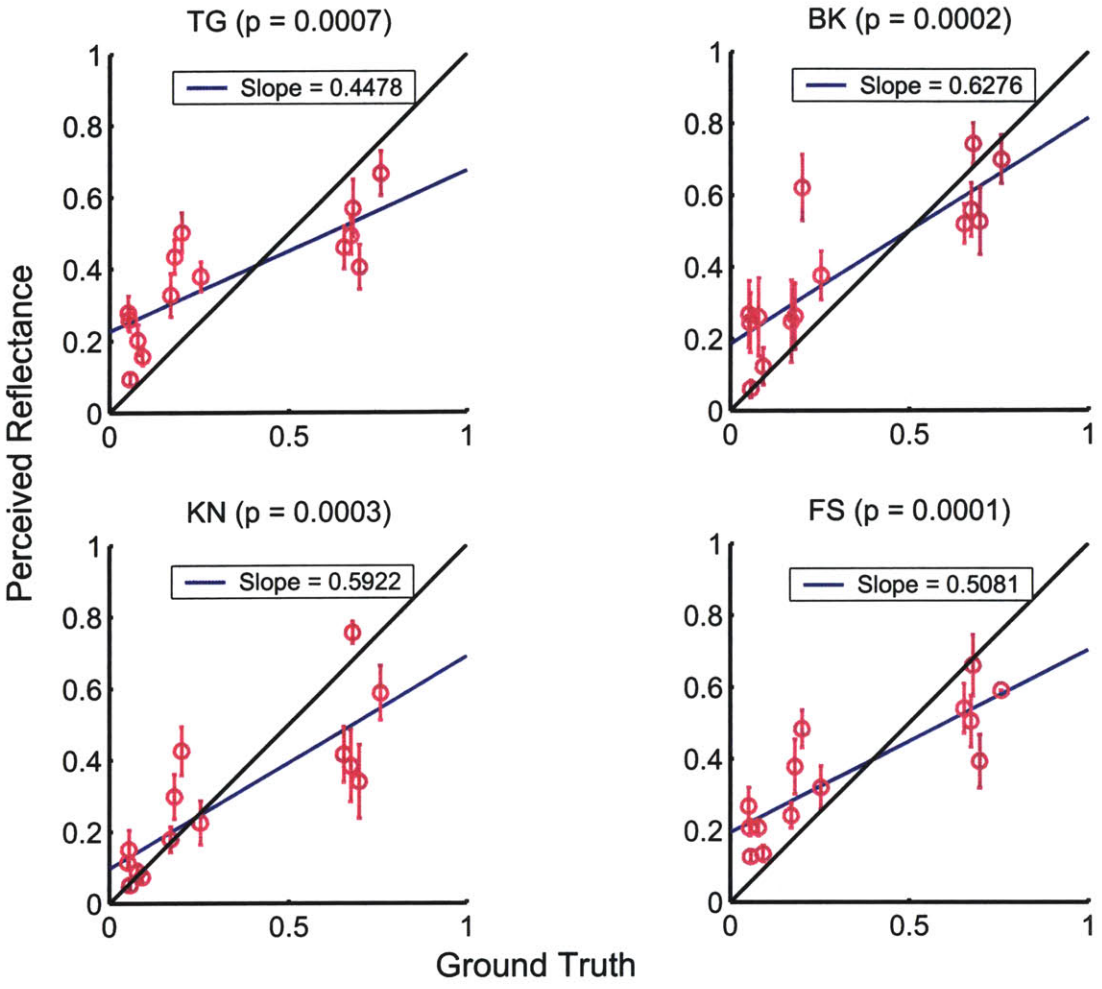


Figure A-30: **Perceived reflectance vs Ground Truth (Group 1)**. The mean responses (pooled over all lighting conditions) are plotted against the ground truth for each observer. Errorbars indicate the 95% confidence intervals. The responses of a veridical observer would lie along the black line with slope = 1 . The blue line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers the slope of the fit is significantly different from 0 and 1.

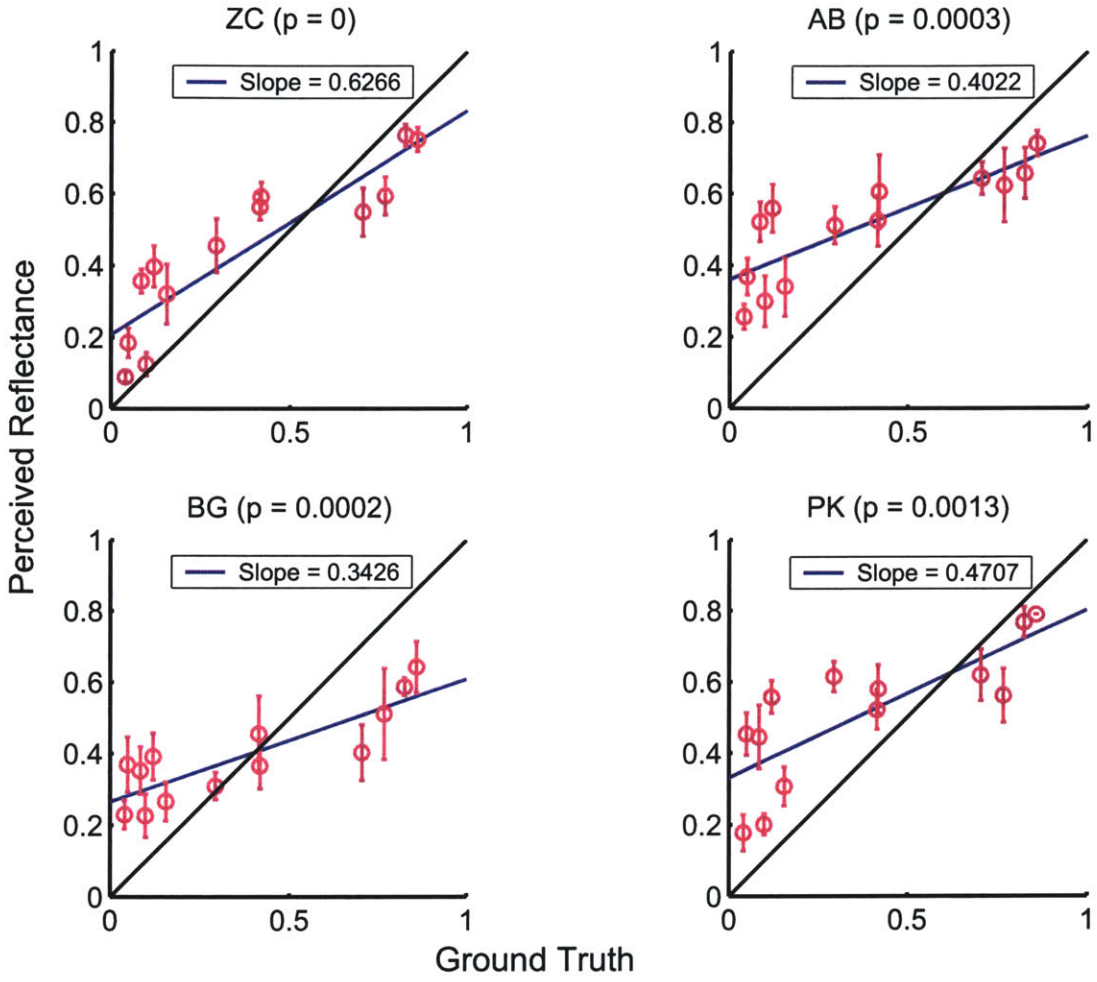


Figure A-31: **Perceived reflectance vs Ground Truth (Group 2).** The mean responses (pooled over all lighting conditions) are plotted against the ground truth for each observer. Errorbars indicate the 95% confidence intervals. The responses of a veridical observer would lie along the black line with slope = 1 . The blue line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For all observers the slope of the fit is significantly different from 0 and 1.

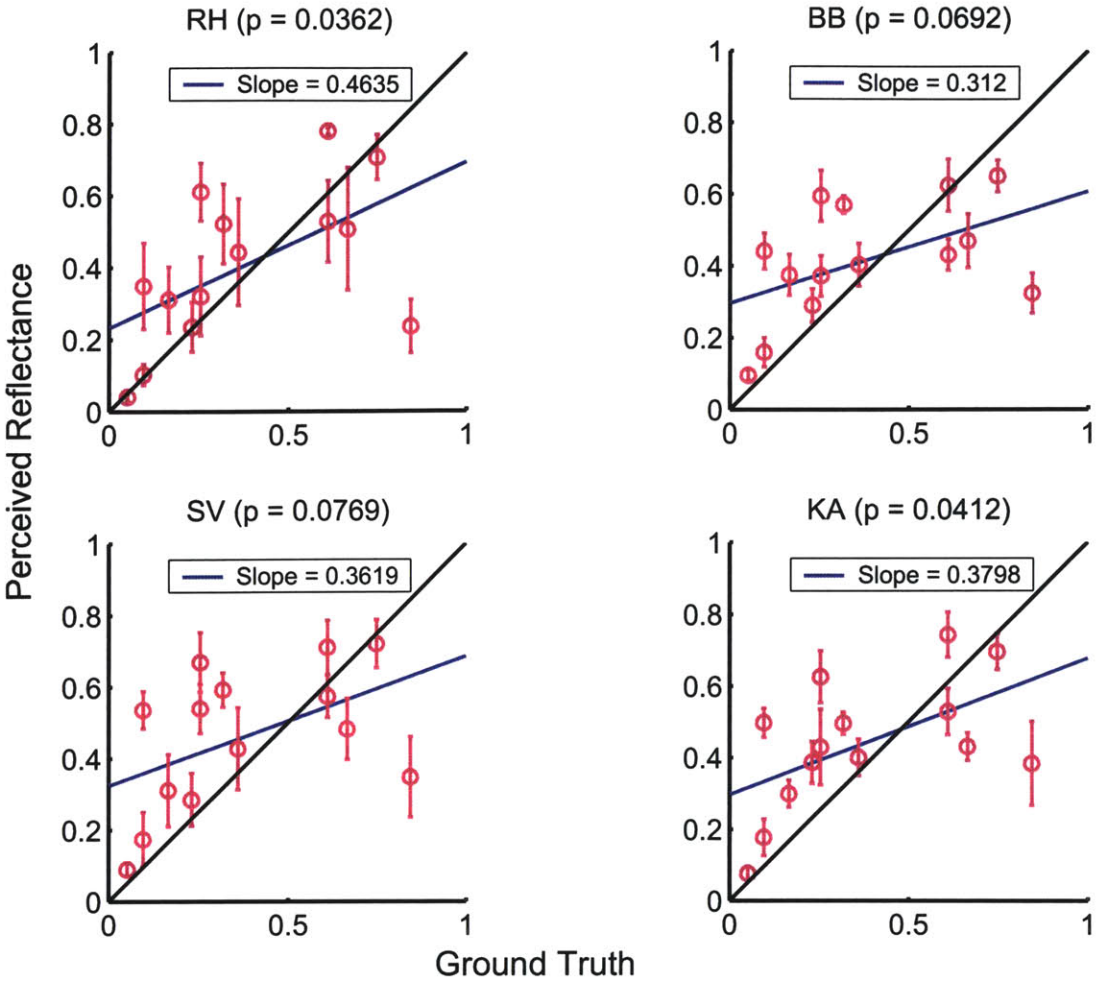


Figure A-32: **Perceived reflectance vs Ground Truth (Group 3)**. The mean responses (pooled over all lighting conditions) are plotted against the ground truth for each observer. Errorbars indicate the 95% confidence intervals. The responses of a veridical observer would lie along the black line with slope = 1 . The blue line is the linear regression fit to each observer's data. The slope of the line and p value are indicated in each plot. For two observers, RH and KA, the slope of the fit is significantly different from 0 and 1. For observers BB and SV, the slope is not significant.

Appendix B

Analysis of Experiment II

B.1 Experiment IIA

In Appendix A, we saw graphical plots of the responses of observers against factors like the mean image luminance and background luminance. We can analyze our data further with ANOVA techniques. Analysis of variance tells us the effect of several factors (in our case observers, lighting, reflectance of material, mean image luminance, background luminance) and their interactions on the measured variable (perceived reflectance). ANOVA assumes a normal distribution and homogeneity of variance if the null hypothesis (factors have no influence on measured variable) is true. ANOVA has been demonstrated to be robust to violations of these assumptions. We perform within-subjects ANOVA for our data. A within subjects design implies that all subjects (observers) view the same stimuli. This is true of all subjects within a group but not across groups. In this analysis, we will not consider across subject factors.

Figures B-1 and B-2 show a three factor within subjects analysis for sections 1,2 and 3 of Experiment IIA. To interpret the ANOVA tables, for each row we consider the value in the last column ($Prob > F$), if it is less than 0.05, then the factor in the first column has an effect on the perceived reflectance. For example for Figure B-1a we observe that all the factors Subject, Lighting, Reflectance and Background Luminance have an effect on the responses of observers in Section 1. Therefore, the

observers do not agree too much with each. Observer responses are affected by the lighting conditions. The responses vary with the reflectance of the material (eg R channel has higher rating than B channel). Background luminance has an effect on the responses. Thus, the ANOVA confirms our observations from Appendix A.

A drawback of ANOVA is that is that it is hard to tell which effect is more significant than others. For example, by running more observers we may reduce the effect of the individual differences between observers. Nevertheless we expect the ANOVA to reflect all the big effects that we observe from graphical plotting of data.

ANOVA tables for Experiment IIA (Figures B-1 through B-3) demonstrate that the mean image luminance, background luminance and the reflectance of the material affect the perceived reflectance of observers. Observers display individual differences and are affected by the lighting conditions.

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.0554	3	0.01846	3.14	0.0266
Lighting	0.2558	2	0.12788	21.76	0
Reflectance	8.7931	2	4.39655	748.05	0
Bkgnd Lum	1.4239	4	0.35598	60.57	0
Subject*Lighting	0.0326	6	0.00544	0.92	0.4784
Subject*Reflectance	0.0586	6	0.00977	1.66	0.1326
Subject*Bkgnd Lum	0.1537	12	0.01281	2.18	0.0144
Lighting*Reflectance	0.082	4	0.0205	3.49	0.009
Lighting*Bkgnd Lum	0.0939	8	0.01174	2	0.0491
Reflectance*Bkgnd Lum	1.2625	8	0.15781	26.85	0
Subject*Lighting*Reflectance	0.049	12	0.00408	0.69	0.7558
Subject*Lighting*Bkgnd Lum	0.1507	24	0.00628	1.07	0.3843
Subject*Reflectance*Bkgnd Lum	0.2124	24	0.00885	1.51	0.07
Lighting*Reflectance*Bkgnd Lum	0.1459	16	0.00912	1.55	0.0864
Subject*Lighting*Reflectance*Bkgnd Lum	0.2355	48	0.00491	0.83	0.7662
Error	1.0579	180	0.00588		
Total	14.0629	359			

Constrained (Type III) sums of squares.

(a) Image Luminance fixed

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.0651	3	0.02172	3.89	0.01
Lighting	0.0839	2	0.04195	7.52	0.0007
Reflectance	7.9332	2	3.96658	710.77	0
Mean Lum	1.781	4	0.44525	79.78	0
Subject*Lighting	0.0141	6	0.00235	0.42	0.8647
Subject*Reflectance	0.2422	6	0.04036	7.23	0
Subject*Mean Lum	0.0939	12	0.00783	1.4	0.1681
Lighting*Reflectance	0.0368	4	0.0092	1.65	0.1641
Lighting*Mean Lum	0.027	8	0.00338	0.61	0.7727
Reflectance*Mean Lum	0.7186	8	0.08983	16.1	0
Subject*Lighting*Reflectance	0.1445	12	0.01204	2.16	0.0155
Subject*Lighting*Mean Lum	0.1229	24	0.00512	0.92	0.5778
Subject*Reflectance*Mean Lum	0.3337	24	0.01391	2.49	0.0003
Lighting*Reflectance*Mean Lum	0.0996	16	0.00623	1.12	0.3433
Subject*Lighting*Reflectance*Mean Lum	0.3049	48	0.00635	1.14	0.2698
Error	1.0045	180	0.00558		
Total	13.0061	359			

Constrained (Type III) sums of squares.

(b) Background Luminance fixed

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.0651	3	0.02172	3.89	0.01
Lighting	0.0839	2	0.04195	7.52	0.0007
Reflectance	7.9332	2	3.96658	710.77	0
Mean Lum	1.781	4	0.44525	79.78	0
Subject*Lighting	0.0141	6	0.00235	0.42	0.8647
Subject*Reflectance	0.2422	6	0.04036	7.23	0
Subject*Mean Lum	0.0939	12	0.00783	1.4	0.1681
Lighting*Reflectance	0.0368	4	0.0092	1.65	0.1641
Lighting*Mean Lum	0.027	8	0.00338	0.61	0.7727
Reflectance*Mean Lum	0.7186	8	0.08983	16.1	0
Subject*Lighting*Reflectance	0.1445	12	0.01204	2.16	0.0155
Subject*Lighting*Mean Lum	0.1229	24	0.00512	0.92	0.5778
Subject*Reflectance*Mean Lum	0.3337	24	0.01391	2.49	0.0003
Lighting*Reflectance*Mean Lum	0.0996	16	0.00623	1.12	0.3433
Subject*Lighting*Reflectance*Mean Lum	0.3049	48	0.00635	1.14	0.2698
Error	1.0045	180	0.00558		
Total	13.0061	359			

Constrained (Type III) sums of squares.

(c) Ratio of image to background luminance fixed

Figure B-1: Three Factor Within Subjects ANOVA for Material 10

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.96976	3	0.32325	37.76	0
Lighting	0.18195	2	0.09098	10.63	0
Reflectance	1.15304	2	0.57652	67.34	0
Bkgnd Lum	1.68097	4	0.42024	49.09	0
Subject*Lighting	0.06733	6	0.01122	1.31	0.2545
Subject*Reflectance	0.06645	6	0.01108	1.29	0.2622
Subject*Bkgnd Lum	0.61157	12	0.05096	5.95	0
Lighting*Reflectance	0.00715	4	0.00179	0.21	0.9333
Lighting*Bkgnd Lum	0.16041	8	0.02005	2.34	0.0204
Reflectance*Bkgnd Lum	0.4029	8	0.05036	5.88	0
Subject*Lighting*Reflectance	0.1066	12	0.00888	1.04	0.4164
Subject*Lighting*Bkgnd Lum	0.25944	24	0.01081	1.26	0.1957
Subject*Reflectance*Bkgnd Lum	0.25838	24	0.01077	1.26	0.1996
Lighting*Reflectance*Bkgnd Lum	0.17605	16	0.011	1.29	0.2106
Subject*Lighting*Reflectance*Bkgnd Lum	0.47225	48	0.00984	1.15	0.2559
Error	1.54101	180	0.00856		
Total	8.11525	359			

Constrained (Type III) sums of squares.

(a) Image Luminance fixed

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	1.34259	3	0.44753	81.12	0
Lighting	0.20538	2	0.10269	18.61	0
Reflectance	0.97684	2	0.48842	88.54	0
Mean Lum	2.84947	4	0.71237	129.13	0
Subject*Lighting	0.03374	6	0.00562	1.02	0.4142
Subject*Reflectance	0.08101	6	0.0135	2.45	0.0267
Subject*Mean Lum	0.45528	12	0.03794	6.88	0
Lighting*Reflectance	0.0218	4	0.00545	0.99	0.4156
Lighting*Mean Lum	0.09297	8	0.01162	2.11	0.0373
Reflectance*Mean Lum	0.29624	8	0.03703	6.71	0
Subject*Lighting*Reflectance	0.11036	12	0.0092	1.67	0.0774
Subject*Lighting*Mean Lum	0.17522	24	0.0073	1.32	0.154
Subject*Reflectance*Mean Lum	0.17749	24	0.0074	1.34	0.1436
Lighting*Reflectance*Mean Lum	0.0274	16	0.00171	0.31	0.9953
Subject*Lighting*Reflectance*Mean Lum	0.14881	48	0.0031	0.56	0.9897
Error	0.993	180	0.00552		
Total	7.9876	359			

Constrained (Type III) sums of squares.

(b) Background Luminance fixed

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	1.34259	3	0.44753	81.12	0
Lighting	0.20538	2	0.10269	18.61	0
Reflectance	0.97684	2	0.48842	88.54	0
Mean Lum	2.84947	4	0.71237	129.13	0
Subject*Lighting	0.03374	6	0.00562	1.02	0.4142
Subject*Reflectance	0.08101	6	0.0135	2.45	0.0267
Subject*Mean Lum	0.45528	12	0.03794	6.88	0
Lighting*Reflectance	0.0218	4	0.00545	0.99	0.4156
Lighting*Mean Lum	0.09297	8	0.01162	2.11	0.0373
Reflectance*Mean Lum	0.29624	8	0.03703	6.71	0
Subject*Lighting*Reflectance	0.11036	12	0.0092	1.67	0.0774
Subject*Lighting*Mean Lum	0.17522	24	0.0073	1.32	0.154
Subject*Reflectance*Mean Lum	0.17749	24	0.0074	1.34	0.1436
Lighting*Reflectance*Mean Lum	0.0274	16	0.00171	0.31	0.9953
Subject*Lighting*Reflectance*Mean Lum	0.14881	48	0.0031	0.56	0.9897
Error	0.993	180	0.00552		
Total	7.9876	359			

Constrained (Type III) sums of squares.

(c) Ratio of image to background luminance fixed

Figure B-2: Three Factor Within Subjects ANOVA for Material 5

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	3.4274	3	1.14247	125.95	0
Lighting	0.5922	2	0.2961	32.64	0
Reflectance	3.0943	2	1.54717	170.57	0
Mean Lum	2.9388	5	0.58776	64.8	0
Bkgnd Lum	1.681	9	0.18677	20.59	0
Error	9.5966	1058	0.00907		
Total	24.1719	1079			

Constrained (Type III) sums of squares.

(a) Material 5

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.1607	3	0.0536	5.73	0.0007
Lighting	0.389	2	0.1945	20.8	0
Reflectance	24.6445	2	12.3222	1317.53	0
Mean Lum	1.7951	5	0.359	38.39	0
Bkgnd Lum	1.4239	9	0.1582	16.92	0
Error	9.895	1058	0.0094		
Total	40.0797	1079			

Constrained (Type III) sums of squares.

(b) Material 10

Figure B-3: Four Factor Within Subjects ANOVA for Materials 5 and 10

B.2 Experiment IIC

ANOVA tables for this task (Figure B-4) indicate that observers display individual differences and that lightness judgements are affected by the reflectance of the material. Except for Group 1, the lighting condition affects observer responses.

To analyze the performance of the observers in this experiment further, Figures B-5 through B-13 plot the absolute value of the difference between perceived reflectance and ground truth for each (material, channel, light, observer) combination for each group. Such a visualization allows one to see how much and on which materials observers deviate from the ground truth. It also facilitates a comparison between observers. For Groups 1 and 3 observers agree with each other and make similar mistakes on the same materials (ANOVA tests confirm this observation). For Group 2 however, observers display individual differences in their errors. The second observation we make from these plots is that error depends on the material i.e. some materials are harder to judge than others (ANOVA tests confirm this observation).

In Figures B-14, B-15 and B-16, the absolute error is plotted against the reflectance of the material for each light and each group. From the plots, we observe the error does not seem to be related to the reflectance of the material. In other words the absolute deviation from ground truth does not depend on the ground truth. ANOVA tests however differ from this observation.

Finally, we examine the effect of manipulating image statistics. In Figures B-17 through B-23, the mean response for each version (R,B2R,G,R2B,B) for each orange material is graphed as a bar plot for each group in each light. These plots allows us to examine the success of image statistics at changing the perception of a material. Image R and B2R have identical histograms and filter output histograms. Therefore according to our chosen set of statistics they are indistinguishable. If these statistics capture anything of perceptual relevance then images R and B2R should be rated

identically by all observers. The same reasoning holds for the B and R2B images. From the plots we find that for nearly all materials and under all lights, both the (R,B2R) and (B,R2B) mean response pairs are within 2 standard error bars of each other. This is a very satisfying result as it confirms that our chosen statistics capture perceptually relevant image information.

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.7286	3	0.24288	27.28	0
Lighting	0.0315	2	0.01575	1.77	0.1721
Reflectance	16.079	13	1.23685	138.91	0
Subject*Lighting	0.0581	6	0.00969	1.09	0.3689
Subject*Reflectance	0.9675	39	0.02481	2.79	0
Lighting*Reflectance	0.2295	26	0.00883	0.99	0.4789
Subject*Lighting*Reflectance	0.7684	78	0.00985	1.11	0.2706
Error	2.9917	336	0.0089		
Total	21.8544	503			

Constrained (Type III) sums of squares.

(a) Group 1

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	1.1212	3	0.37374	64.27	0
Lighting	0.3158	2	0.15789	27.15	0
Reflectance	11.9302	12	0.99418	170.95	0
Subject*Lighting	0.0578	6	0.00964	1.66	0.1311
Subject*Reflectance	1.4842	36	0.04123	7.09	0
Lighting*Reflectance	0.8097	24	0.03374	5.8	0
Subject*Lighting*Reflectance	0.7914	72	0.01099	1.89	0.0001
Error	1.8145	312	0.00582		
Total	18.3247	467			

Constrained (Type III) sums of squares.

(b) Group 2

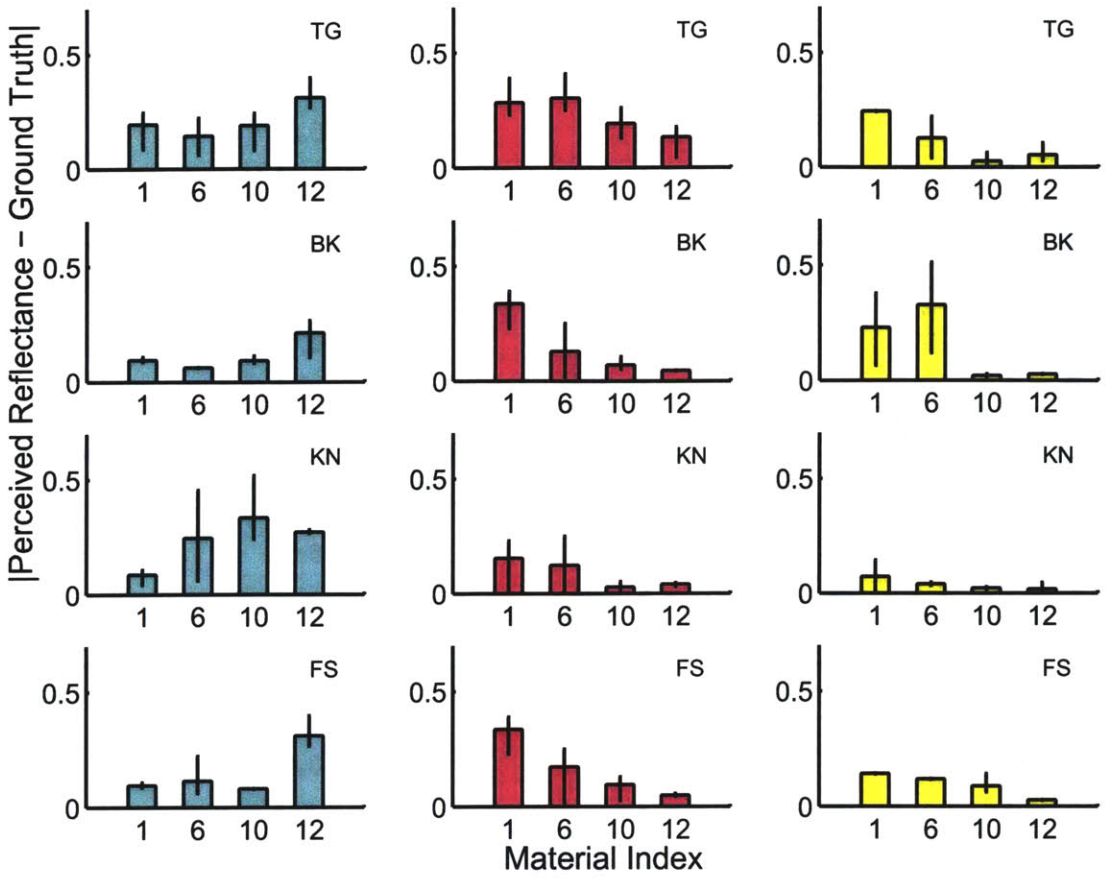
Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.1381	3	0.04603	2.76	0.0418
Lighting	0.7493	2	0.37465	22.5	0
Reflectance	13.1613	10	1.31613	79.03	0
Subject*Lighting	0.5497	6	0.09162	5.5	0
Subject*Reflectance	0.7116	30	0.02372	1.42	0.0723
Lighting*Reflectance	0.9166	20	0.04583	2.75	0.0001
Subject*Lighting*Reflectance	0.8517	60	0.0142	0.85	0.7725
Error	6.1948	372	0.01665		
Total	23.2363	503			

Constrained (Type III) sums of squares.

(c) Group 3

Figure B-4: Two Factor Within Subjects ANOVA, Keeping Background and Image Luminance Fixed



(a)

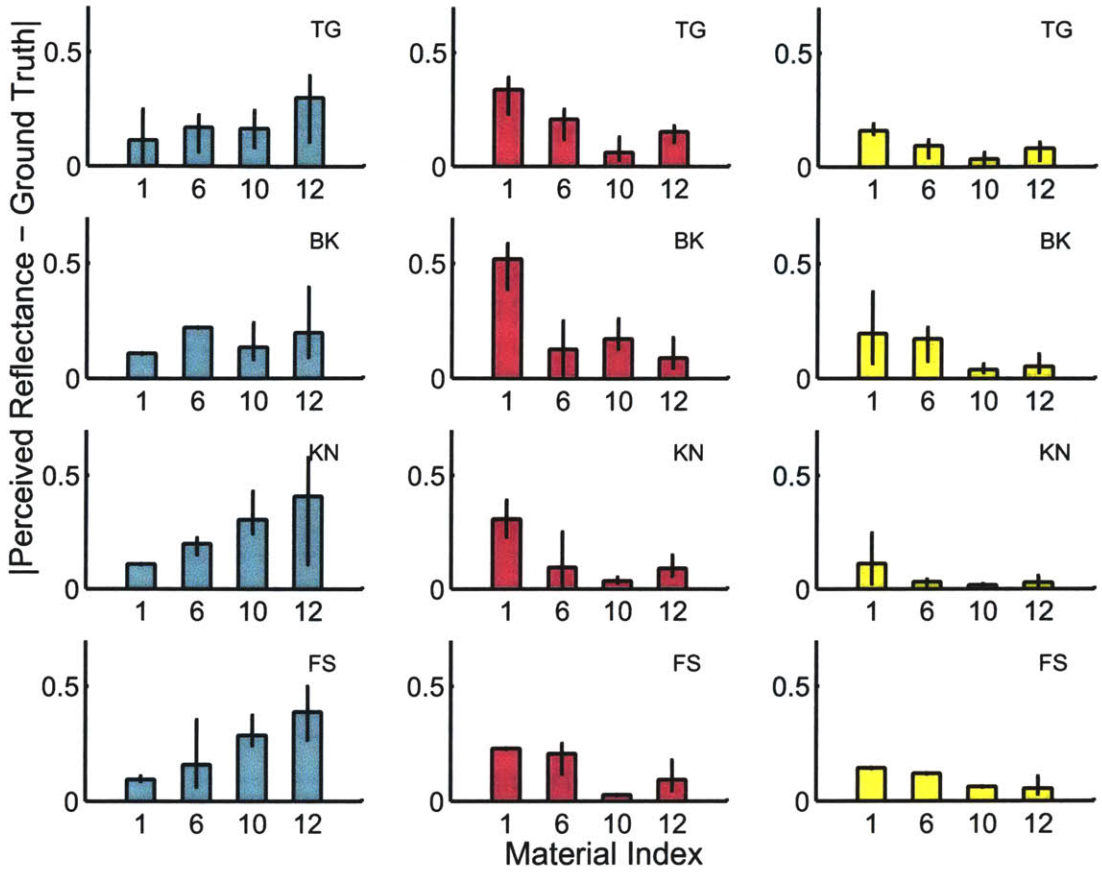
Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.04408	3	0.01469	1.11	0.3466
Material	0.19887	5	0.03977	3.01	0.013
Subject*Material	0.26719	15	0.01781	1.35	0.1819
Error	1.90411	144	0.01322		
Total	2.44375	167			

Constrained (Type III) sums of squares.

(b)

Figure B-5: (a) **|Perceived reflectance - Ground Truth| vs Material Index (Group 1, Light 1)**. The mean absolute difference between ground truth and perceived reflectance is plotted against the material for each observer. Errorbars indicate the range. Each row corresponds to a different observer and each column to the R (cyan), G (magenta) or B (yellow) channels for each material. Note how most observers agree with each other and how some materials are harder to judge than others. (b) One Factor Within Subjects ANOVA, (Image and Background Luminance are fixed) confirms these observations.



(a)

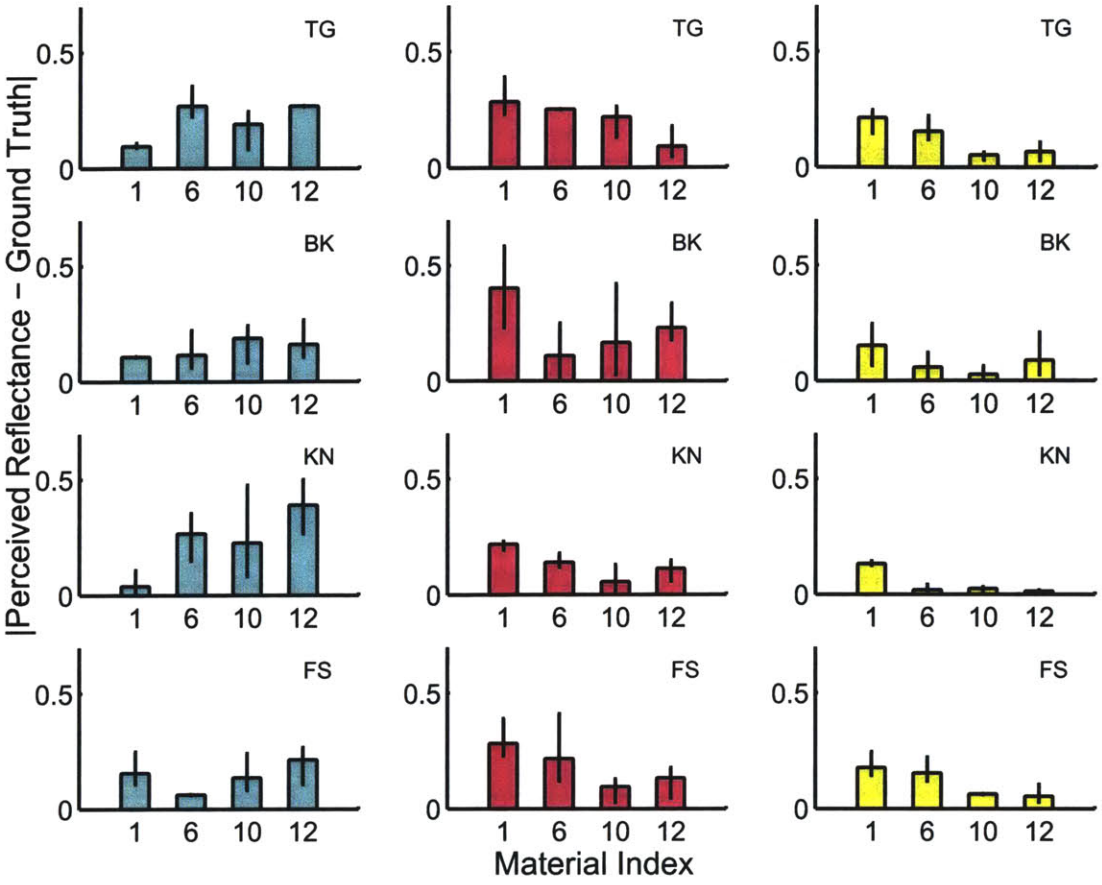
Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.02162	3	0.00721	0.45	0.7154
Material	0.16473	5	0.03295	2.07	0.0723
Subject*Material	0.15755	15	0.0105	0.66	0.8193
Error	2.28994	144	0.0159		
Total	2.62707	167			

Constrained (Type III) sums of squares.

(b)

Figure B-6: (a) **|Perceived reflectance - Ground Truth| vs Material Index (Group 1, Light 2)** (b) One Factor Within Subjects ANOVA (Image and Background Luminance are fixed)



(a)

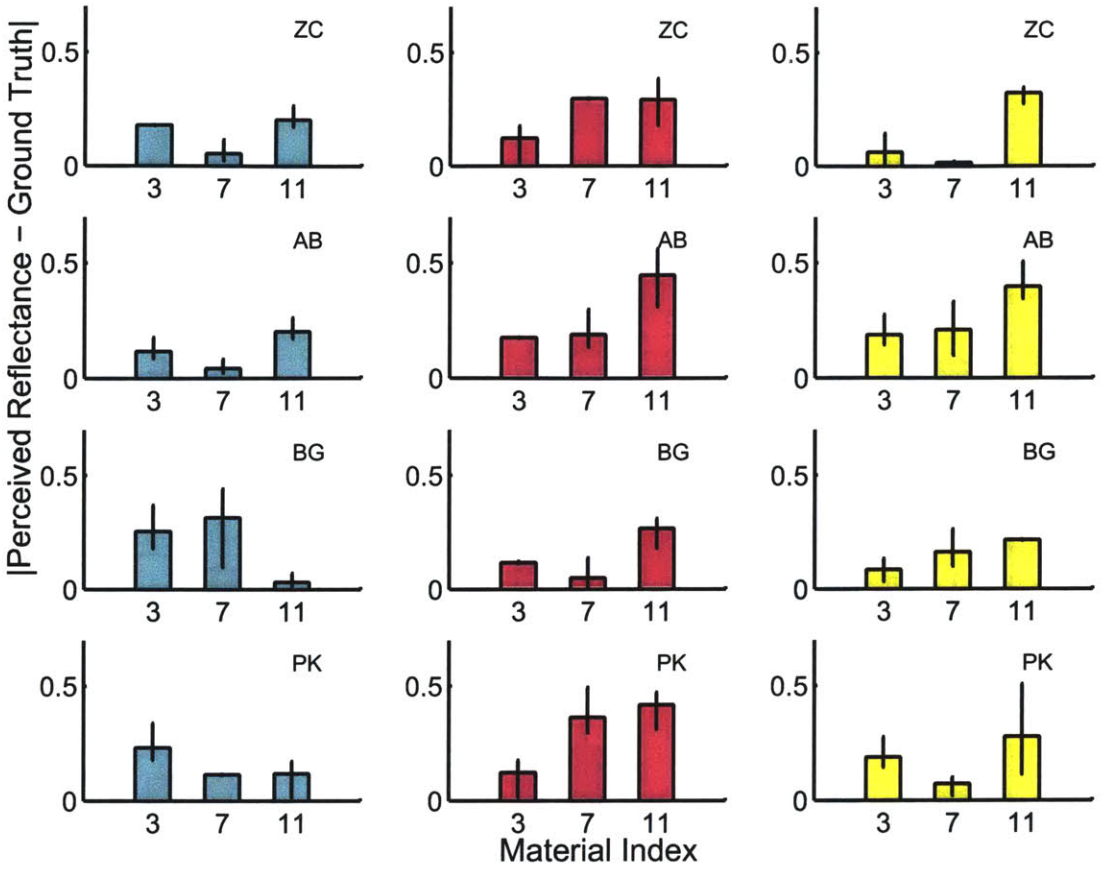
Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.04533	3	0.01511	1.21	0.3081
Material	0.10573	5	0.02115	1.69	0.1396
Subject*Material	0.1735	15	0.01157	0.93	0.5362
Error	1.79723	144	0.01248		
Total	2.12003	167			

Constrained (Type III) sums of squares.

(b)

Figure B-7: (a) **|Perceived reflectance - Ground Truth| vs Material Index (Group 1, Light 3)** (b) One Factor Within Subjects ANOVA (Image and Background Luminance are fixed)



(a)

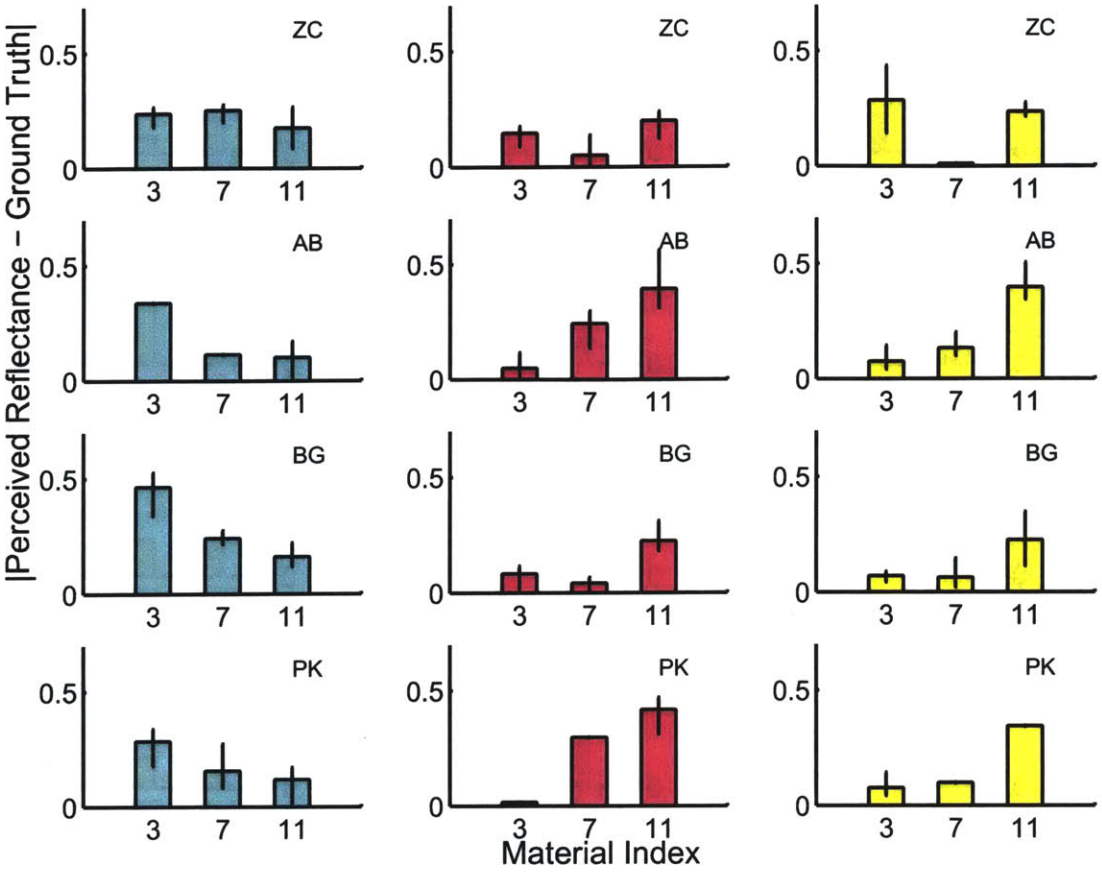
Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.1199	3	0.03997	3.44	0.0189
Material	0.34957	6	0.05826	5.01	0.0001
Subject*Material	0.43136	18	0.02396	2.06	0.0107
Error	1.4872	128	0.01162		
Total	2.34137	155			

Constrained (Type III) sums of squares.

(b)

Figure B-8: (a) **|Perceived reflectance - Ground Truth| vs Material Index (Group 2, Light 1)** (b) One Factor Within Subjects ANOVA (Image and Background Luminance are fixed)



(a)

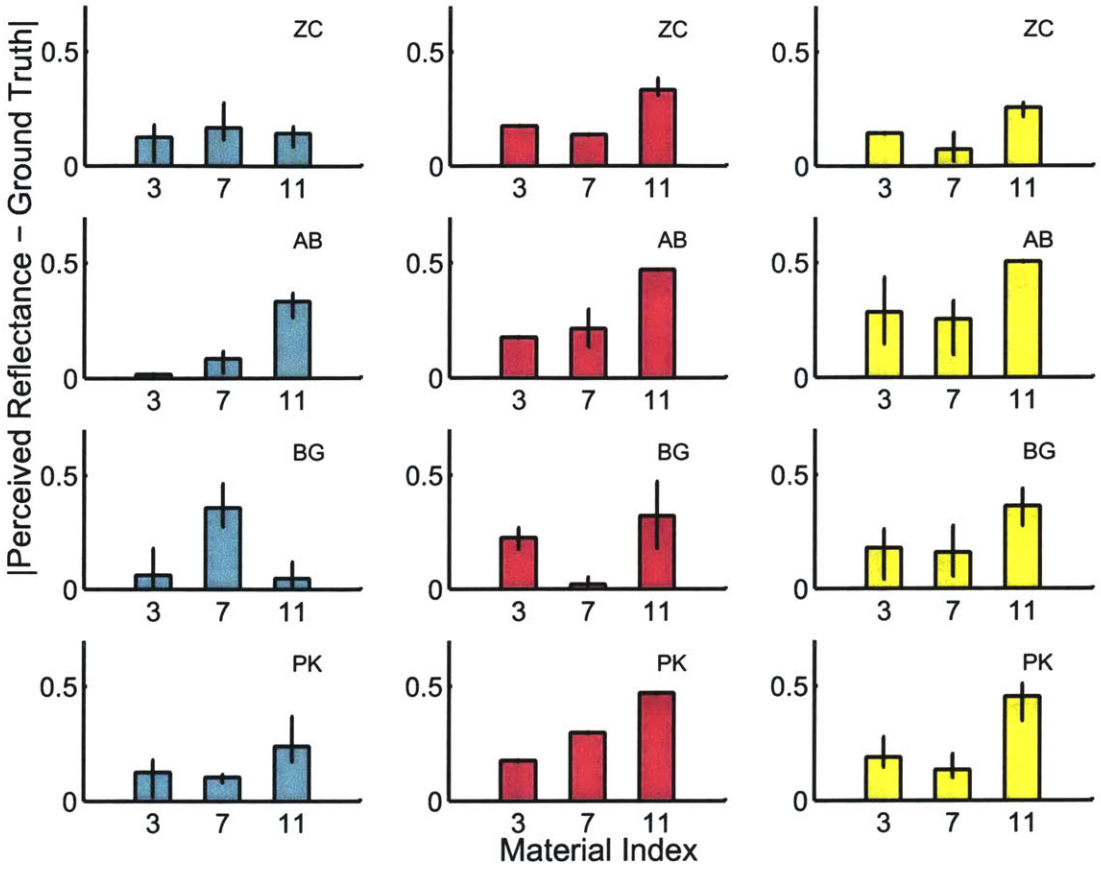
Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.17391	3	0.05797	4.39	0.0056
Material	0.73502	6	0.1225	9.28	0
Subject*Material	0.47107	18	0.02617	1.98	0.015
Error	1.68998	128	0.0132		
Total	2.98716	155			

Constrained (Type III) sums of squares.

(b)

Figure B-9: (a) **|Perceived reflectance - Ground Truth| vs Material Index (Group 2, Light 2)** (b) One Factor Within Subjects ANOVA (Image and Background Luminance are fixed)



(a)

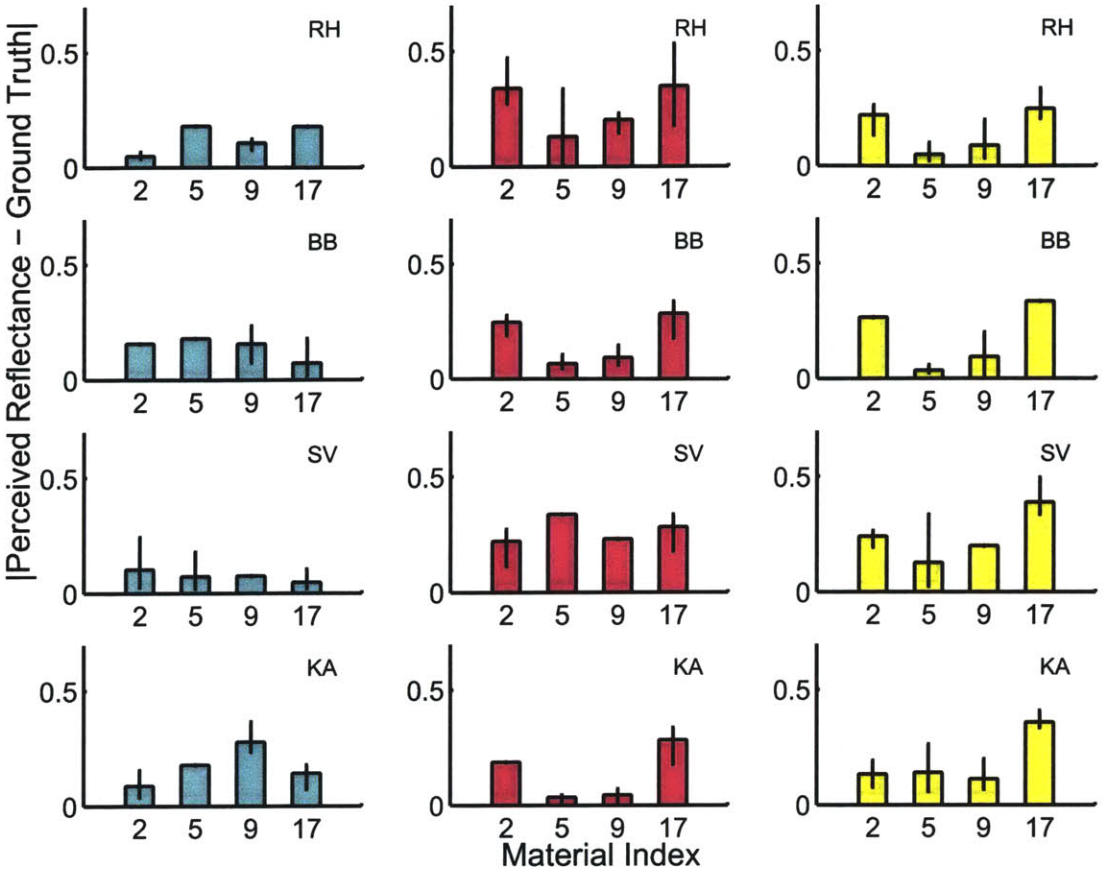
Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.22286	3	0.07429	6.88	0.0002
Material	0.97852	6	0.16309	15.1	0
Subject*Material	0.30319	18	0.01684	1.56	0.0805
Error	1.38243	128	0.0108		
Total	2.88354	155			

Constrained (Type III) sums of squares.

(b)

Figure B-10: (a) **|Perceived reflectance - Ground Truth| vs Material Index (Group 2, Light 3)** (b) One Factor Within Subjects ANOVA (Image and Background Luminance are fixed)



(a)

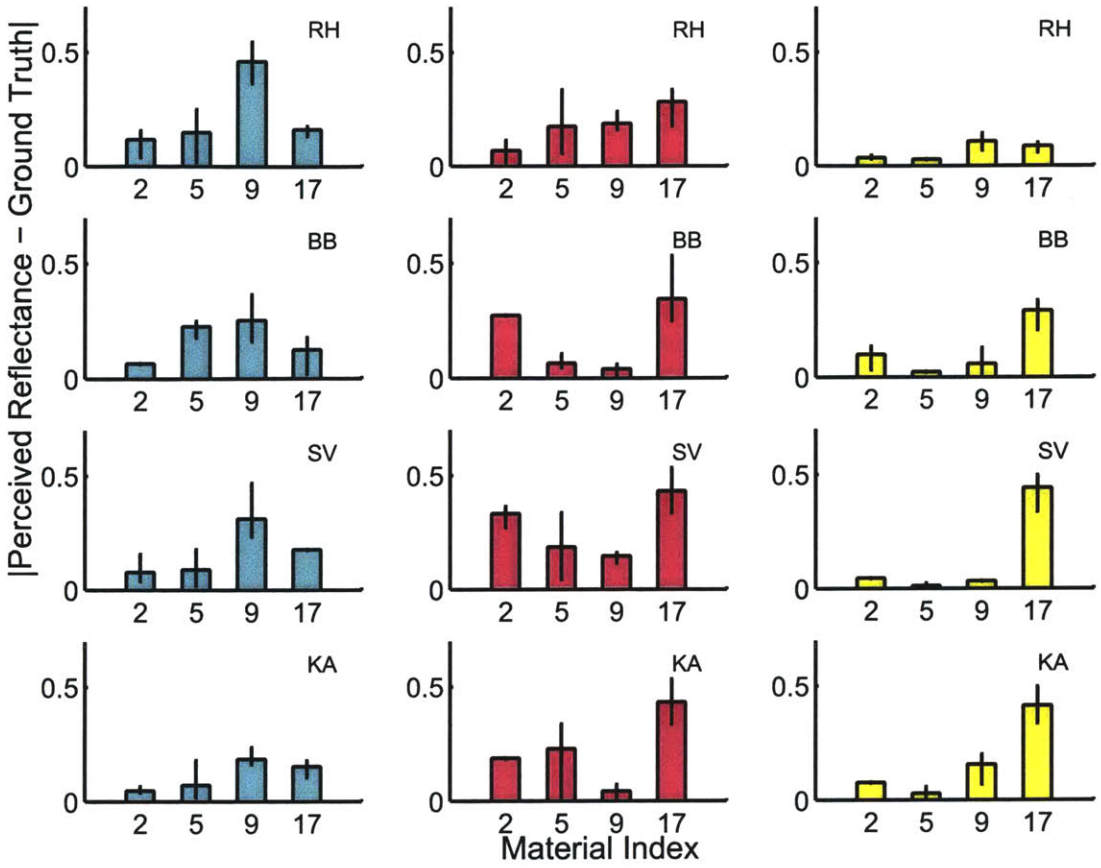
Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.00815	3	0.00272	0.25	0.8602
Material	2.10655	5	0.42131	39.01	0
Subject*Material	0.16328	15	0.01089	1.01	0.4505
Error	1.55528	144	0.0108		
Total	3.83299	167			

Constrained (Type III) sums of squares.

(b)

Figure B-11: (a) **|Perceived reflectance - Ground Truth| vs Material Index (Group 3, Light 1)** (b) One Factor Within Subjects ANOVA (Image and Background Luminance are fixed)



(a)

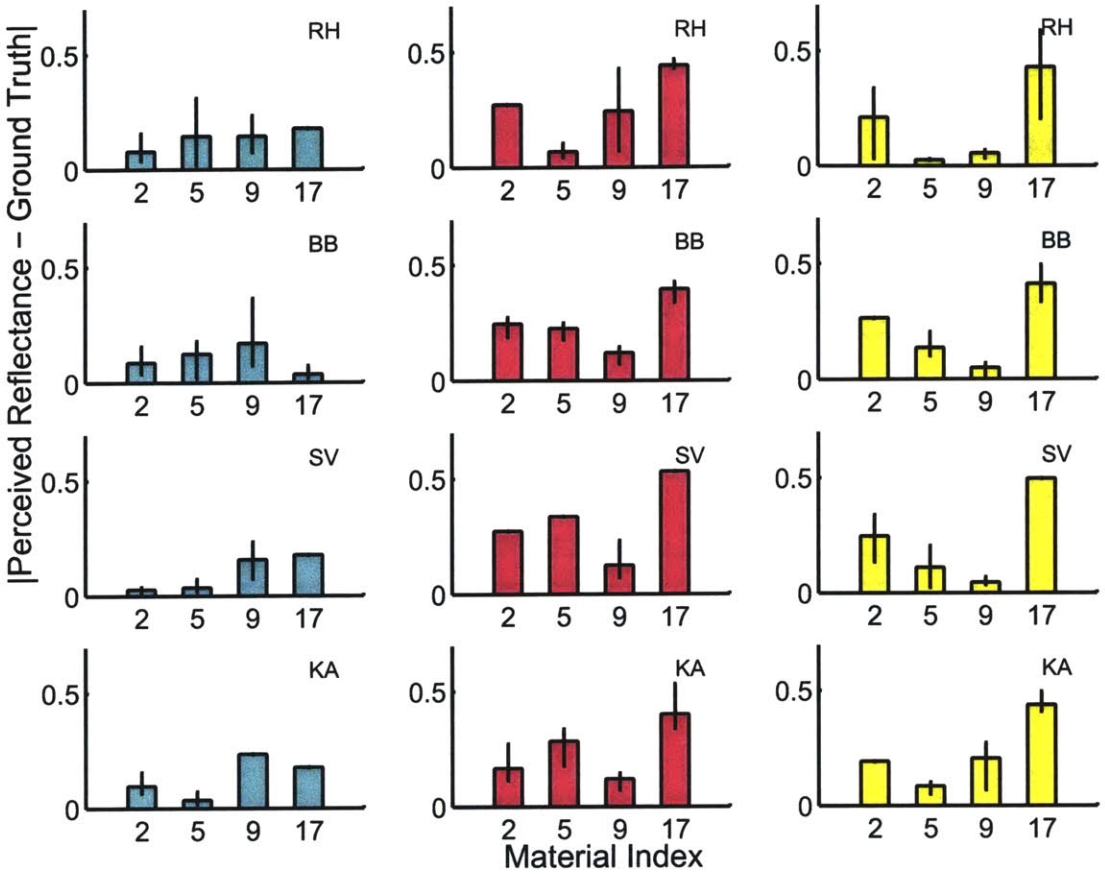
Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.02998	3	0.00999	0.72	0.5403
Material	3.06341	5	0.61268	44.28	0
Subject*Material	0.36845	15	0.02456	1.78	0.0435
Error	1.99264	144	0.01384		
Total	5.4543	167			

Constrained (Type III) sums of squares.

(b)

Figure B-12: (a) **|Perceived reflectance - Ground Truth| vs Material Index (Group 3, Light 2)** (b) One Factor Within Subjects ANOVA (Image and Background Luminance are fixed)



(a)

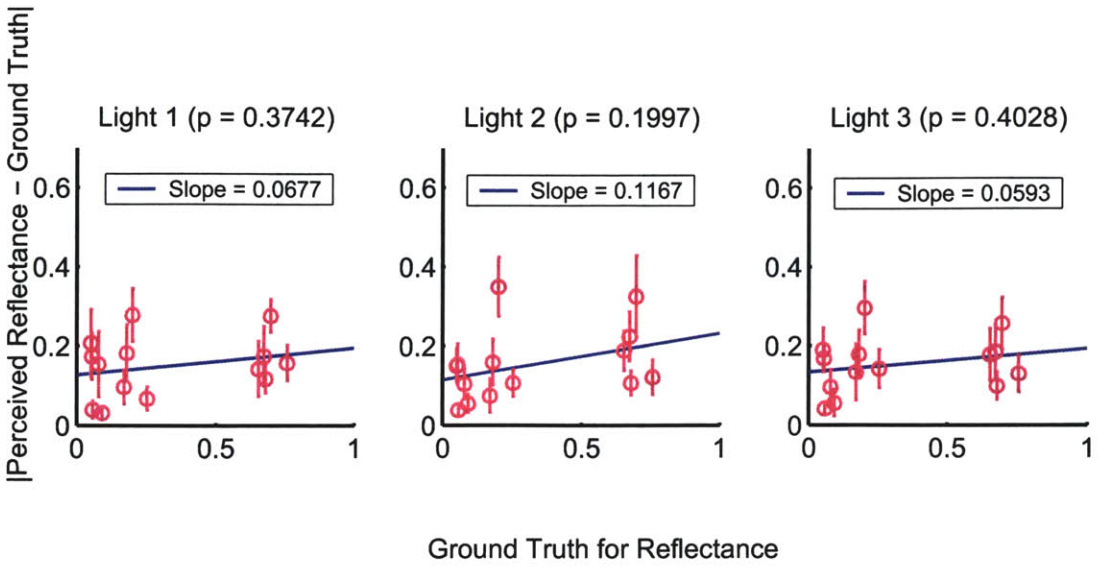
Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.0303	3	0.0101	0.7	0.5535
Material	1.90162	5	0.38032	26.36	0
Subject*Material	0.25814	15	0.01721	1.19	0.2837
Error	2.07767	144	0.01443		
Total	4.2494	167			

Constrained (Type III) sums of squares.

(b)

Figure B-13: (a) **|Perceived reflectance - Ground Truth| vs Material Index (Group 3, Light 3)** (b) One Factor Within Subjects ANOVA (Image and Background Luminance are fixed)



(a)

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.07359	3	0.02453	3.57	0.0163
Reflectance	0.89476	13	0.06883	10.02	0
Subject*Reflectance	0.70627	39	0.01811	2.64	0
Error	0.76914	112	0.00687		
Total	2.44375	167			

Constrained (Type III) sums of squares.

(b)

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.01485	3	0.00495	0.66	0.5798
Reflectance	1.3226	13	0.10174	13.52	0
Subject*Reflectance	0.44682	39	0.01146	1.52	0.0459
Error	0.8428	112	0.00753		
Total	2.62707	167			

Constrained (Type III) sums of squares.

(c)

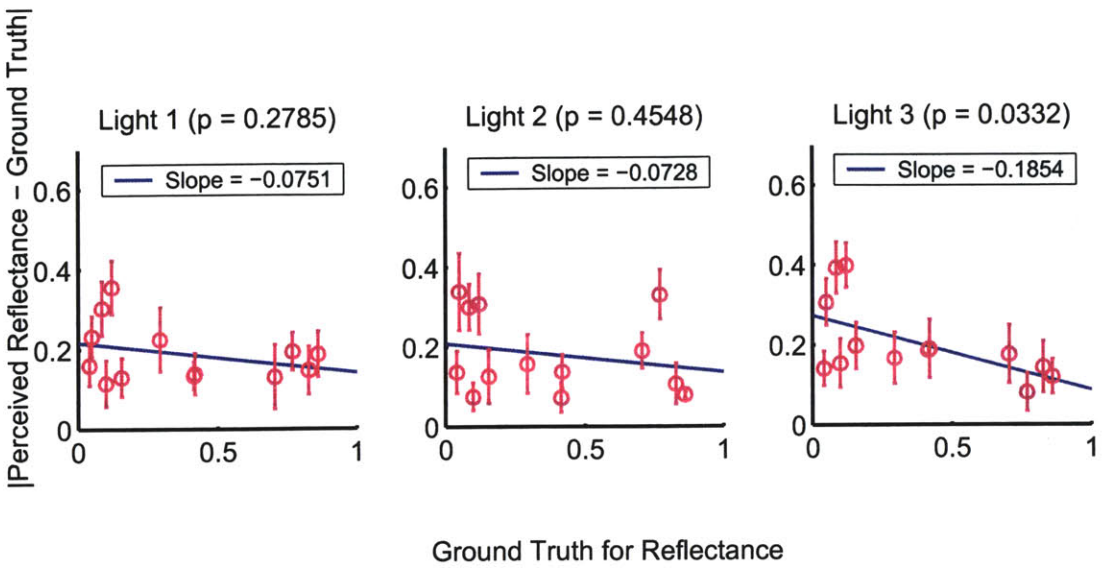
Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.04357	3	0.01452	1.97	0.1227
Reflectance	0.76969	13	0.05921	8.03	0
Subject*Reflectance	0.48083	39	0.01233	1.67	0.0195
Error	0.82594	112	0.00737		
Total	2.12003	167			

Constrained (Type III) sums of squares.

(d)

Figure B-14: (a) **|Perceived reflectance - Ground Truth| vs Ground Truth (Group 1)** (b) One Factor Within Subjects ANOVA (Image and Background Luminance are fixed) Light 1 (c) Light 2 (d) Light 3



(a)

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.07323	3	0.02441	4.22	0.0074
Reflectance	0.77159	12	0.0643	11.12	0
Subject*Reflectance	0.8951	36	0.02486	4.3	0
Error	0.60145	104	0.00578		
Total	2.34137	155			

Constrained (Type III) sums of squares.

(b)

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.09108	3	0.03036	6.42	0.0005
Reflectance	1.48067	12	0.12339	26.09	0
Subject*Reflectance	0.92357	36	0.02565	5.42	0
Error	0.49184	104	0.00473		
Total	2.98716	155			

Constrained (Type III) sums of squares.

(c)

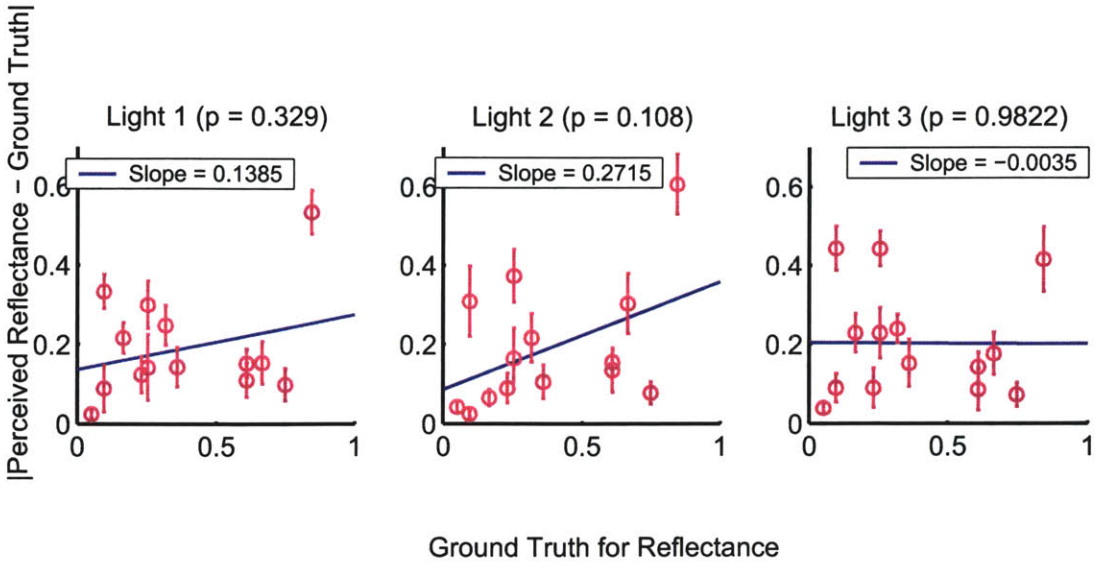
Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.21941	3	0.07314	13.96	0
Reflectance	1.41917	12	0.11826	22.58	0
Subject*Reflectance	0.70021	36	0.01945	3.71	0
Error	0.54474	104	0.00524		
Total	2.88354	155			

Constrained (Type III) sums of squares.

(d)

Figure B-15: (a) **|Perceived reflectance - Ground Truth| vs Ground Truth (Group 2)** (b) One Factor Within Subjects ANOVA (Image and Background Luminance are fixed) Light 1 (c) Light 2 (d) Light 3



(a)

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.00377	3	0.00126	0.13	0.9422
Reflectance	2.11537	10	0.21154	21.84	0
Subject*Reflectance	0.5087	30	0.01696	1.75	0.0177
Error	1.20104	124	0.00969		
Total	3.83299	167			

Constrained (Type III) sums of squares.

(b)

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.03279	3	0.01093	0.85	0.4679
Reflectance	3.25293	10	0.32529	25.37	0
Subject*Reflectance	0.58139	30	0.01938	1.51	0.061
Error	1.59018	124	0.01282		
Total	5.4543	167			

Constrained (Type III) sums of squares.

(c)

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Subject	0.00675	3	0.00225	0.15	0.9292
Reflectance	1.99361	10	0.19936	13.33	0
Subject*Reflectance	0.38907	30	0.01297	0.87	0.6656
Error	1.85474	124	0.01496		
Total	4.2494	167			

Constrained (Type III) sums of squares.

(d)

Figure B-16: (a) **|Perceived reflectance - Ground Truth| vs Ground Truth (Group 3)** (b) One Factor Within Subjects ANOVA (Image and Background Luminance are fixed) Light 1 (c) Light 2 (d) Light 3

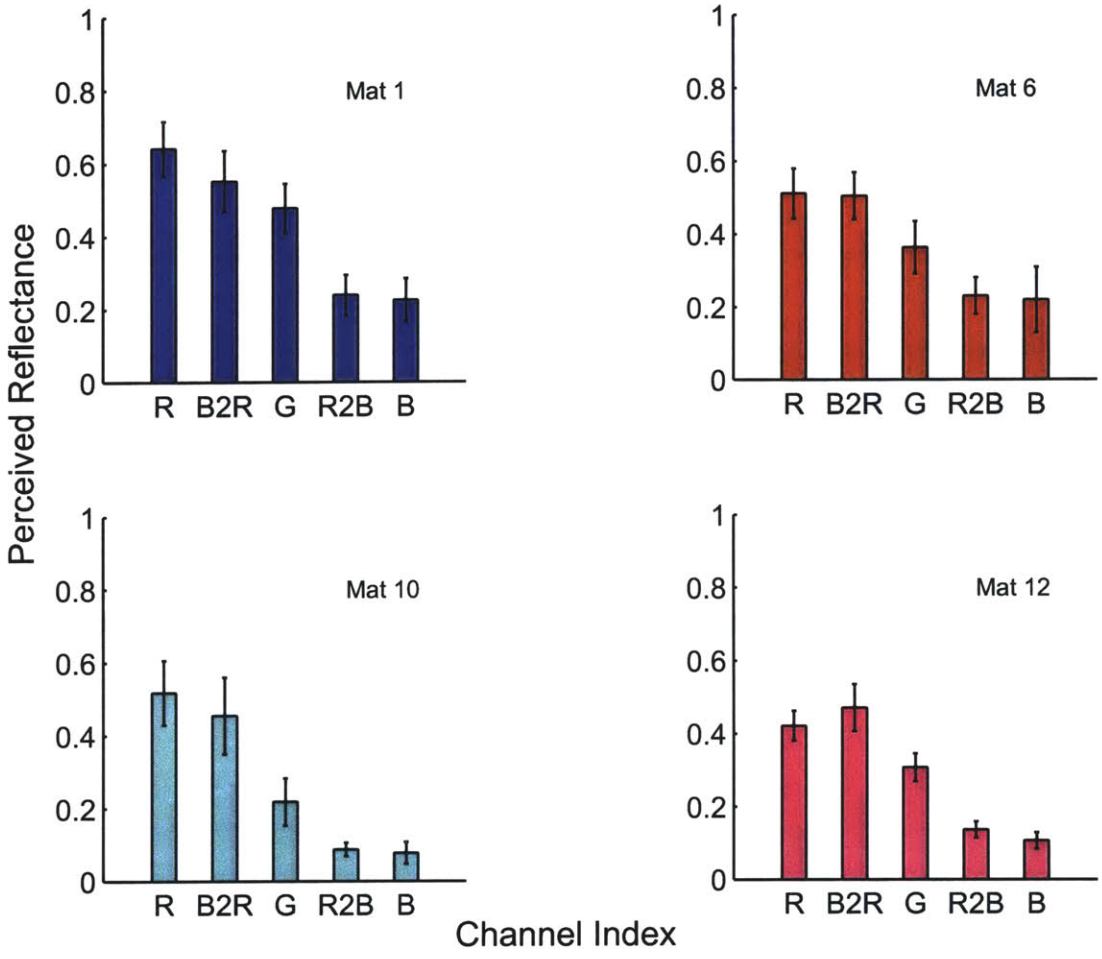


Figure B-17: **Perceived Reflectance vs Channel Index (Group 1, Light 1)** The mean response pooled across subjects is plotted against the channel index for each material. Error bars indicated 95% confidence intervals. The success of manipulated images $R2B$ and $B2R$ varies depending on the material.

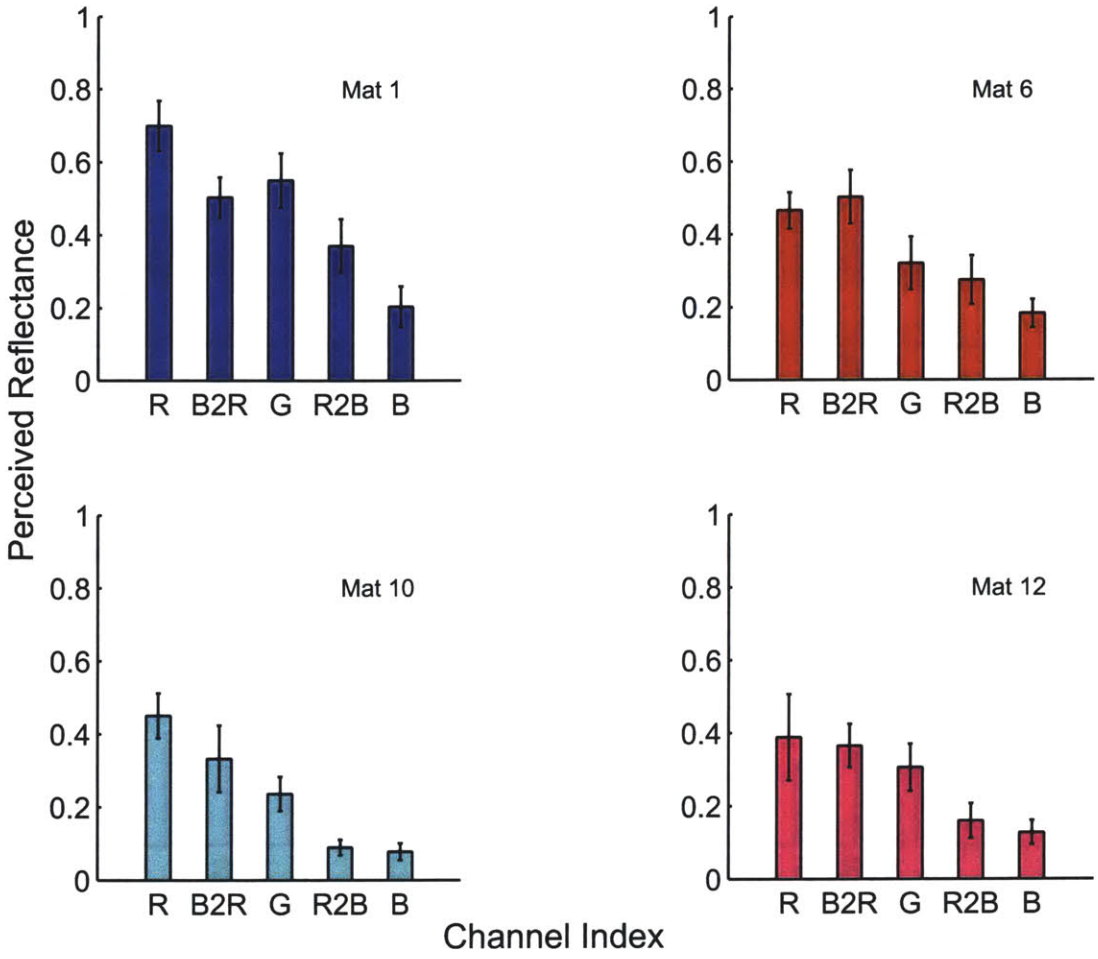


Figure B-18: **Perceived Reflectance vs Channel Index (Group 1, Light 2)**

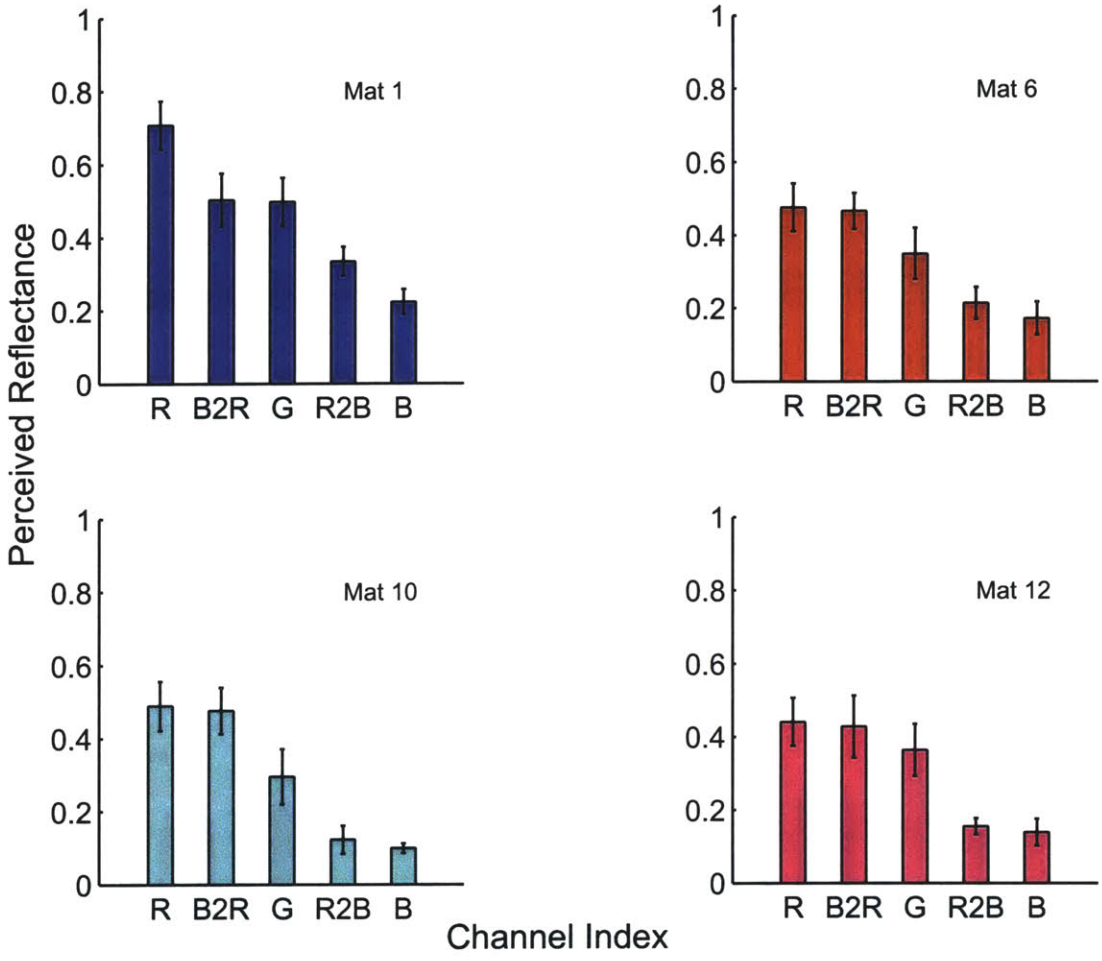
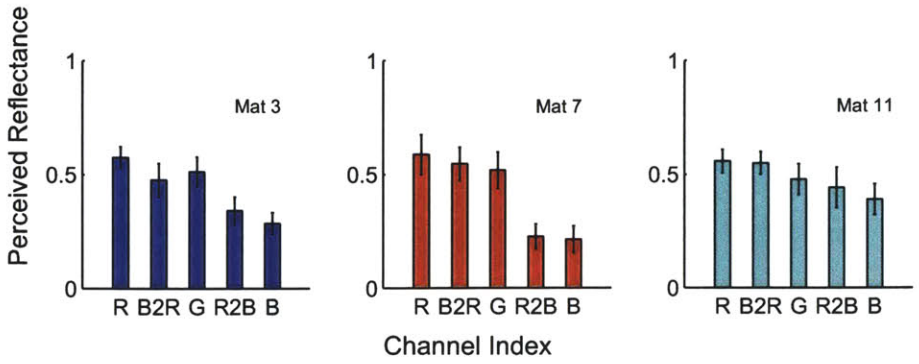
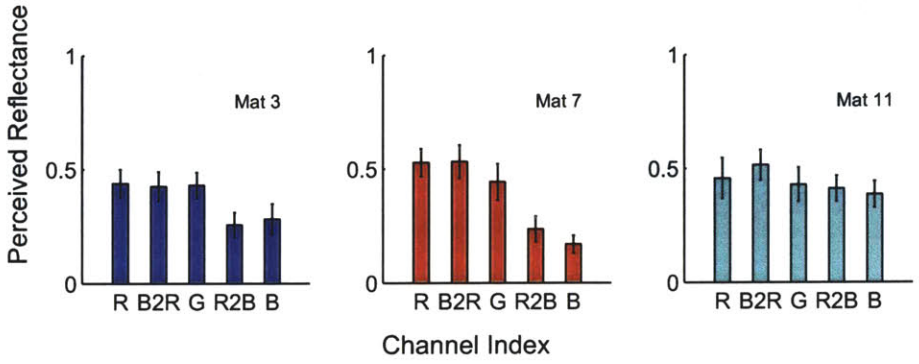


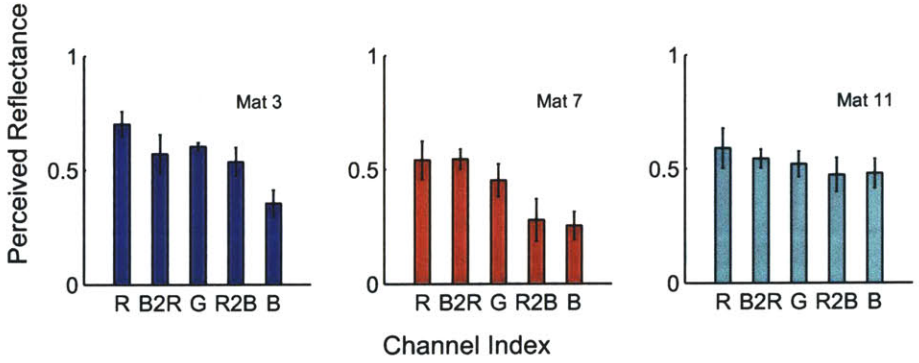
Figure B-19: **Perceived Reflectance vs Channel Index (Group 1, Light 3)**



(a) Light 1



(b) Light 2



(c) Light 3

Figure B-20: **Perceived Reflectance vs Channel Index (Group 2)** The mean response pooled across subjects is plotted against the channel index for each material. Error bars indicated 95% confidence intervals. The success of manipulated images $R2B$ and $B2R$ varies depending on the material.

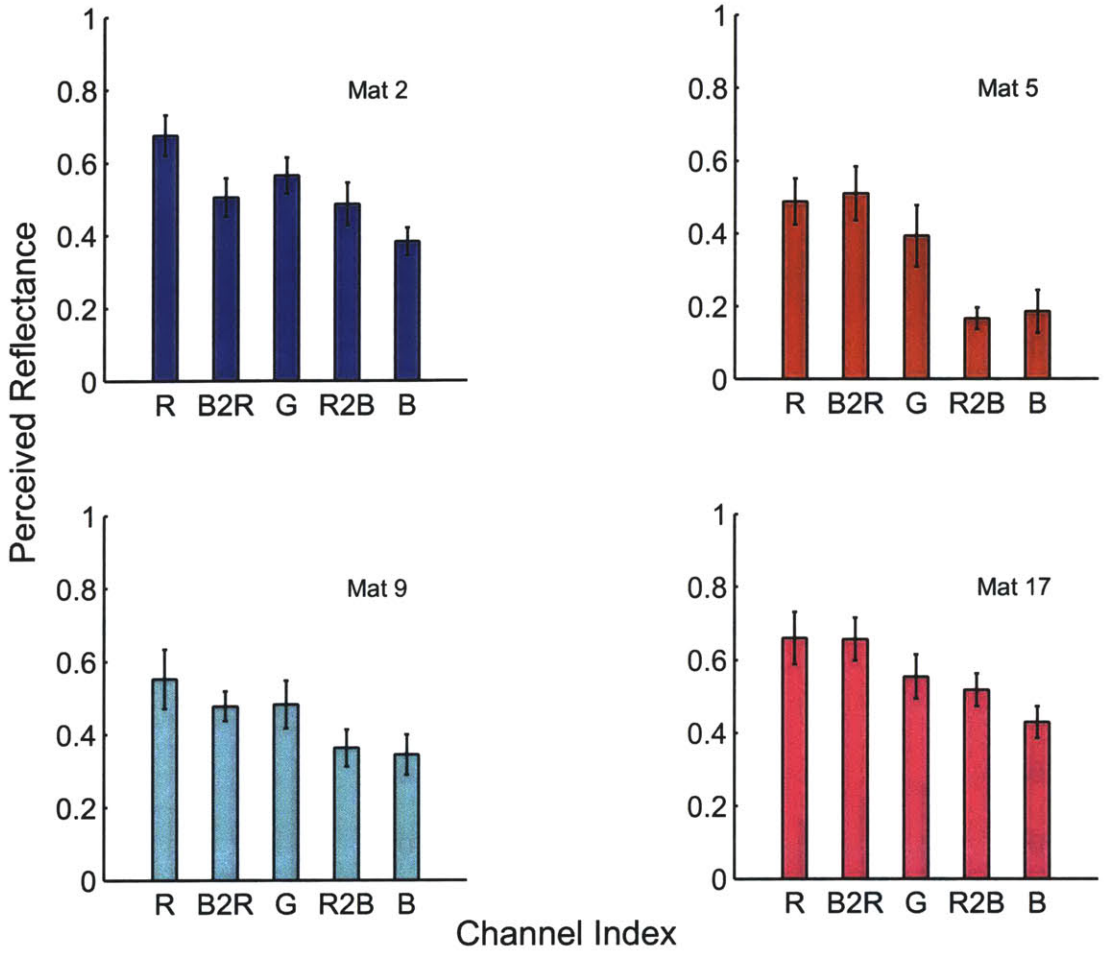


Figure B-21: **Perceived Reflectance vs Channel Index (Group 3, Light 1)** The mean response pooled across subjects is plotted against the channel index for each material. Error bars indicated 95% confidence intervals. The success of manipulated images $R2B$ and $B2R$ varies depending on the material.

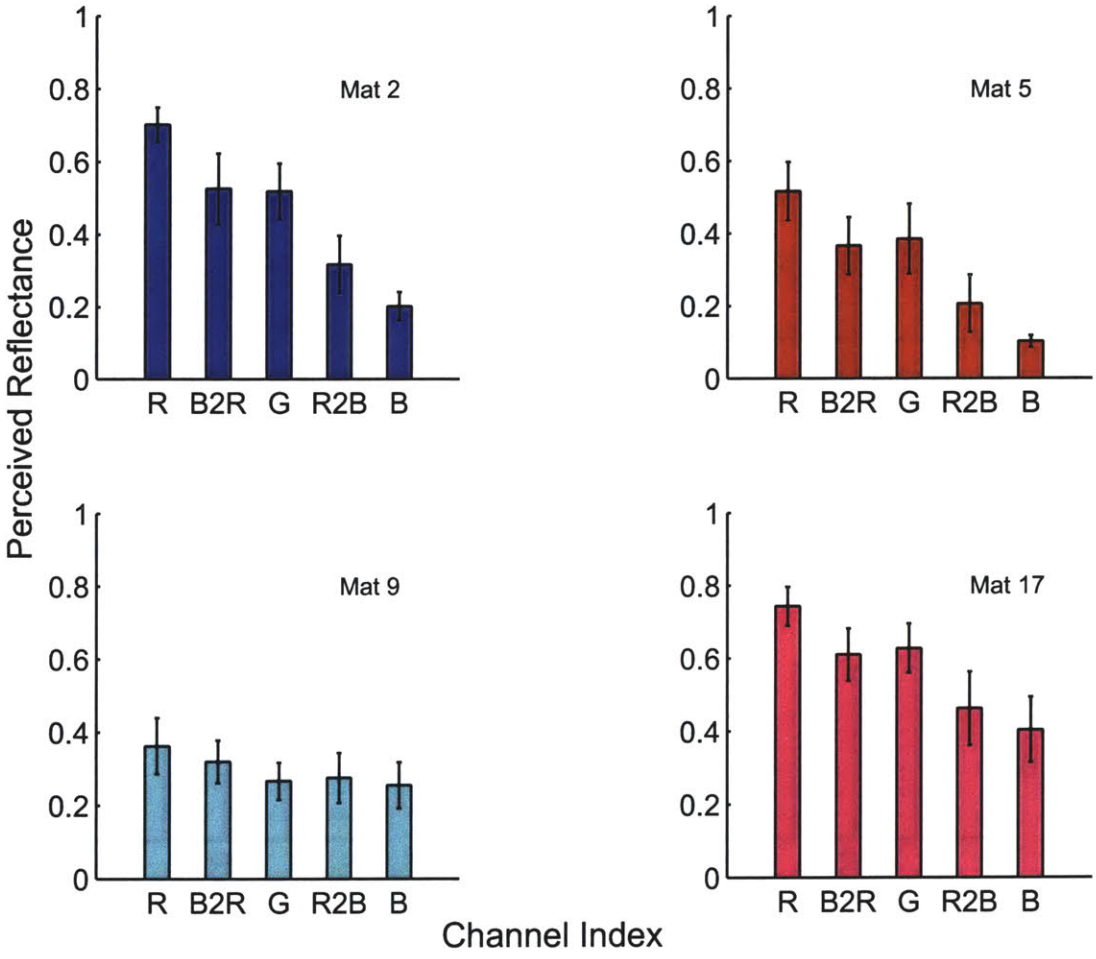


Figure B-22: **Perceived Reflectance vs Channel Index (Group 3, Light 2)**

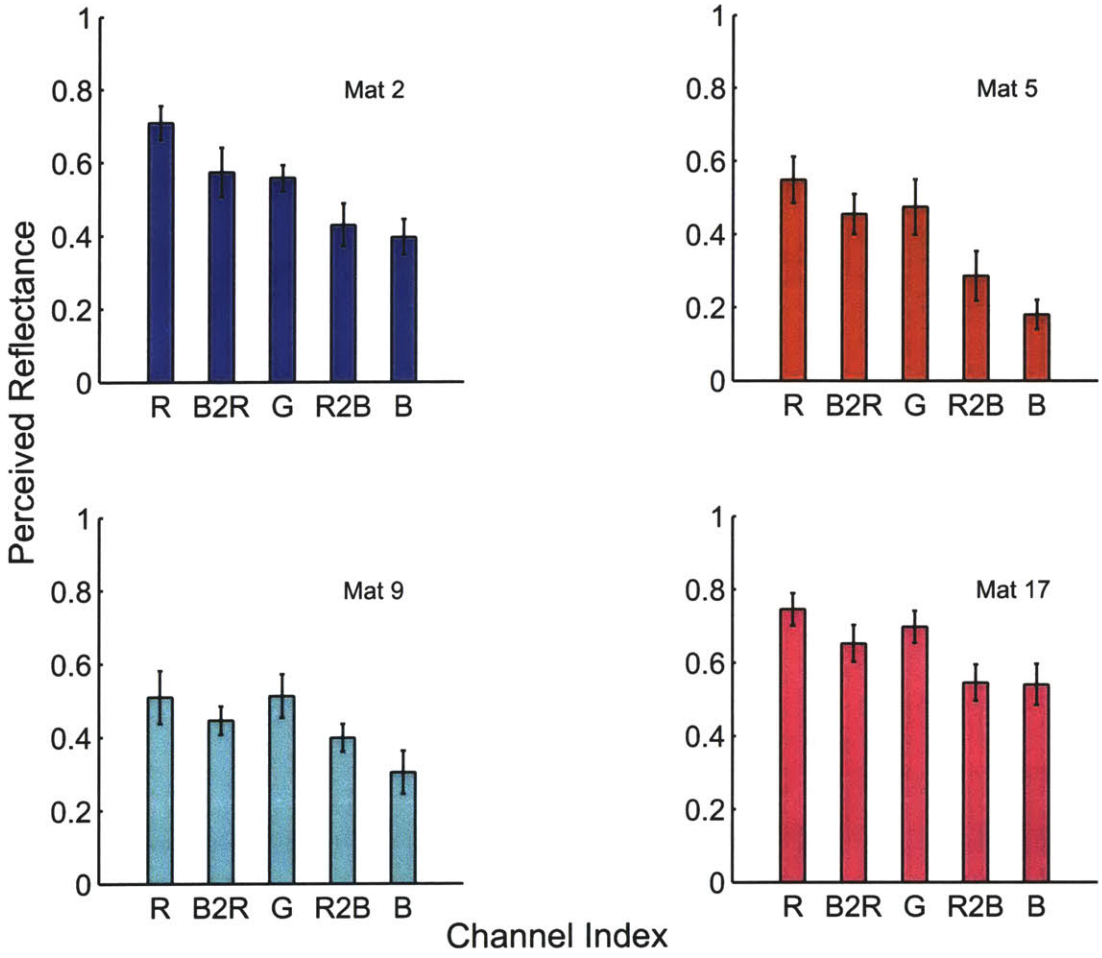


Figure B-23: **Perceived Reflectance vs Channel Index (Group 3, Light 3)**

Bibliography

[1] E. H. Adelson. Lightness perception and lightness illusions. In M. Gazzaniga, editor, *The Cognitive Neurosciences*, pages 339–351. MIT Press, Cambridge MA, 2000.

[2] E. H. Adelson. Textural statistics and surface perception. In *Vision Sciences Society Annual Meeting Abstracts*, May 2003. `http://journalofvision.org/3/9/48/`.

[3] E. H. Adelson, Y. Li, and L. Sharan. Image statistics for material perception. In *Vision Sciences Society Annual Meeting Abstracts*, May 2004. `http://journalofvision.org/5/8/569/`.

[4] J. Beck and S. Prazdny. Highlights and the perception of glossiness. *Perception and Psychophysics*, 30:407–410, 1981.

[5] S. Boivin and A. Gagalowicz. Image based rendering of diffuse, specular and glossy surfaces from a single image. In *SIGGRAPH Conference Proceedings*, 2001.

[6] D. H. Brainard. The psychophysics toolbox. *Spatial Vision*, 10:433–436, 1997.

[7] P. Burt and E. H. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communication*, 31:532–540, 1983.

[8] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[9] Dave Coffin. Dcraw: Raw digital photo decoding in linux. Software available at `http://www.cybercom.net/~dcoffin/dcraw`.

[10] K. J. Dana, B. van Ginneken, S. K. Nayar, and J. J. Koenderink. Reflectance and texture of real world surfaces. *ACM Transactions on Graphics*, 18:1–34, 1999.

[11] P. Debevec, T. Hawkins, C. Tchou, H-P Duiker, W. Sarokin, and M. Sagar. Acquiring the reflectance field of a human face. In *SIGGRAPH Conference Proceedings*, 2000.

[12] P. Debevec, C. Tchou, A. Gardner, T. Hawkins, C. Poullis, J. Stumpfel, A. Jones, N. Yun, P. Einarsson, T. Lundgren, M. Fajardo, and P. Martinez. Estimating surface reflectance properties of a complex scene under captured natural illumination. Technical Report ICT-TR-06, USC ICT, December 2004.

[13] Paul Debevec, Chris Tchou, and Tim Hawkins. *HDRShop: High Dynamic Range Image Processing and Manipulation*, 2004. Software available at `http://www.hdrshop.com`.

[14] R. O. Dror, E. H. Adelson, and A. S. Willsky. Recognition of surface reflectance properties from a single image under unknown real world illumination. In *Workshop on Identifying Objects Across Variation in Lighting at CVPR*, Hawaii, Dec 2001.

[15] R. O. Dror, A. S. Willsky, and E. H. Adelson. Statistical characterization of real world illumination. *Journal of Vision*, 4:821–837, 2004.

[16] Ron O. Dror. *Surface Reflectance Recognition and Real-World Illumination Statistics.* PhD dissertation, Massachusetts Institute of Technology, Department of Electrical Engg and Computer Science, 2002.

[17] A. A. Efros and W. T. Freeman. Image quilting for texture synthesis and transfer. In *SIGGRAPH Conference Proceedings*, 2001.

[18] A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *Proceedings International Conference on Computer Vision*, 1999.

[19] R. W. Fleming, R. O. Dror, and E. H. Adelson. Real-world illumination and the perception of surface reflectance properties. *Journal of Vision*, 3(5):347–368, 2003.

[20] Roland W. Fleming. *Human visual perception under real-world illumination.* PhD dissertation, Massachusetts Institute of Technology, Department of Brain and Cognitive Sciences, 2003.

[21] W. T. Freeman. Exploiting the generic viewpoint assumption. *International Journal of Computer Vision*, 20(3):243, 1996.

[22] W. T. Freeman and E. H. Adelson. Steerable filters for early vision, image analysis and wavelet decomposition. In *International Conference on Computer Vision*, pages 406–415, Osaka, Japan, 1990.

[23] A. Gilchrist, C. Kossyfidis, F. Bonato, T. Agostini, J. Cataliotti, X. Li, B. Spehar, V. Annan, and E. Economou. An anchoring theory of lightness perception. *Pscyhological Review*, 106:795–834, 1999.

[24] X. D. He, K. E. Torrance, F. S. Sillion, and D. P. Greenberg. A comprehensive physical model for light reflection. *Computer Graphics (SIGGRAPH)*, 25(4):175–186, 1991.

[25] D. J. Heeger and J. R. Bergen. Pyramid based texture analysis/synthesis. *Computer Graphics (SIGGRAPH)*, 1995.

[26] H. Helmholtz. *Helmholtz's Treatise on Physiological Optics.* Dover Publications, 1866,1962. Translated from the third German edition by the Optical Society of America in 1924.

[27] E. Hering. *Outlines of a Theory of the Light Sense.* Harvard University Press, Cambridge, MA, 1874,1964.

[28] J. J. Koenderink, A. J. Van Doorn, K. J. Dana, and S. Nayar. Bidirectional reflectance distribution function of thoroughly pitted surfaces. *International Journal of Computer Vision*, 31:129–144, 1999.

[29] J. J. Koenderink and A. J. van Doorn. Illuminance texture due to surface mesostructure. *Journal of the Optical Society of America*, 13(3):452–463, March 1996.

[30] S.R. Marschner, S. H. Westin, E.P.F. Lafortune, K.E. Torrance, and D.P. Greenberg. Image-based BRDF measurement including human skin. In *10th Eurographics Workshop on Rendering*, pages 139–152, June 1999.

[31] I. Motoyoshi, S. Nishida, and E. H. Adelson. Adaptation to skewed image statistics alters perception of glossiness and lightness. In August, editor, *European Conference on Visual Perception*, 2005. `http://www.perceptionweb.com/ecvp05/0168.html`.

[32] I. Motoyoshi, S. Nishida, and E. H. Adelson. Image statistics as a determinant of reflectance estimation. In *Vision Sciences Society Annual Meeting Abstracts*, May 2005. `http://journalofvision.org/5/8/569/`.

[33] I. Motoyoshi, S. Nishida, and E. H. Adelson. Luminance re-mapping for the control of apparent material. In August, editor, *Second Symposium on Applied Perception in Graphics and Visualization*, 2005.

[34] S. Nishida and M. Shinya. Use of image-based information in judgements of surface reflectance. *Journal of the Optical Society of America A*, 15:2951–2965, 1998.

[35] K. Nishino, Z. Zhang, and K. Ikeuchi. Determining reflectance parameters and illumination distributions from a sparse set of images for view-dependent image synthesis. In *Proceeding International Conference on Computer Vision*, pages 599–601, 2001.

[36] M. Oren and S. K. Nayar. Generalization of the lambertian model and implications for machine vision. *International Journal of Computer Vision*, 14(3):227–251, April 1995.

[37] F. Pellacini, J. A. Ferwerda, and D. P. Greenberg. Toward a psychophysically-based light reflection model for image synthesis. In *SIGGRAPH Conference Proceedings*, 2000.

[38] D. G. Pelli. The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10:437–442, 1997.

[39] B-T. Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317, 1975.

[40] S. C. Pont and J. J. Koenderink. Bidirectional texture contrast function. *International Journal of Computer Vision*, 62(1):17–34, 2005.

[41] J. Portilla and E. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 40:49–71, 2000.

[42] R. Ramamoorthi and P. Hanrahan. A signal processing framework for inverse rendering. In *SIGGRAPH Conference Proceedings*, 2001.

[43] R. Robilotto and Q. Zaidi. Limits of lightness identification for real objects under natural viewing conditions. *Journal of Vision*, 4(9), 2004.

[44] Y. Sato, M. D. Wheeler, and K. Ikeuchi. Object shape and reflectance modeling from observation. In *SIGGRAPH Conference Proceedings*, 1997.

[45] L. Sharan, Y. Li, and E. H. Adelson. Image statistics and reflectance estimation. In *Vision Sciences Society Annual Meeting Abstracts*, May 2005. `http://journalofvision.org/5/8/375/`.

[46] E. P. Simoncelli and E. H. Adelson. *Subband Image Coding*, chapter Subband Transforms. Kluwer Academic Publishers, Norwell, MA, 1990.

[47] Eero Simoncelli. matlabpyrtools: Matlab source code for multi-scale image processing. Software available at `http://www.cns.nyu.edu/~lcv/software.html`.

[48] P. Sinha and E. H. Adelson. Recovering reflectance in a world of painted polyhedra. In *Proceedings International Conference on Computer Vision*, pages 156–163, 1993.

[49] M. F. Tappen, W. T. Freeman, and E. H. Adelson. Recovering intrinsic images from a single image. In *Neural Information Processing Systems*, pages 1343–1350, 2002.

[50] M. F. Tappen, W. T. Freeman, and E. H. Adelson. Recovering intrinsic images from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(9):1459–1472, 2005.

[51] D. Todorovic. Lightness and junctions. *Perception*, 26(4):379–394, 1997.

[52] S. Tominaga and N. Tanaka. Estimating reflection parameters from a single color image. *IEEE Computer Graphics and Applications*, 20:58–66, Sept/Oct 2000.

[53] H. Wallach. Brightness constancy and the nature of achromatic colors. *Journal of Experimental Psychology*, 38:310–324, 1948.

[54] G. J. Ward. Measuring and modeling anisotropic reflection. *Computer Graphics (SIGGRAPH)*, 26(2):265–272, 1992.

[55] Y. Weiss. Deriving intrinsic images from image sequences. In *Proceedings International Conference on Computer Vision*, 2001.

[56] S. Yantis, editor. *Visual Perception: Essential Readings*. Psychology Press, 2000.

[57] Y. Yu, P. Debevec, J. Malik, and T. Hawkins. Inverse global illumination: Recovering reflectance models of real scenes from photographs. In *SIGGRAPH Conference Proceedings*, pages 215–24, 1999.

[58] Y. Yu and J. Malik. Recovering photometric properties of architectural scenes from photographs. In *SIGGRAPH Conference Proceedings*, pages 207–217, 1998.

[59] Q. Zaidi, B. Spehar, and M. Shy. Induced effects of backgrounds and foregrounds in three-dimensional configurations: the role of t-junctions. *Perception*, 26:395–408, 1997.

[60] S. C. Zhu, Y. N. Wu, and D. B. Mumford. Filters, random field and maximum entropy: Towards a unified theory of texture modeling. *International Journal of Computer Vision*, 27(2):1–20, 1998.