

Modelling Texture Appearance of Gonioapparent Objects

Saori Kitaguchi

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy in Colour Science

University of Leeds
Department of Colour Science

March 2008

The candidate confirms that the work submitted is his/her own and that appropriate credit has
been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no
quotation from the thesis may be published without proper acknowledgement.

Acknowledgements

I would like to express my deepest gratitude to my supervisor Prof. Ronnier Luo for his continuous guidance, support and encouragement. I would also like to thank my second supervisor Prof. Stephen Westland. This research work would not have been possible without his support and valuable advice.

Thanks also go to Akzo Nobel and the National Physical Laboratory for supporting this project in terms of financial and technological donations. I would particularly like to thank Dr Eric JJ Kirchner and ing Geert-Jan van den Kieboom MSc of Akzo Novel and Prof. Michael R Pointer, former member of NPL and current member of the Department of Colour Science, the University of Leeds.

Special thanks go to Prof. Tetsuya Sato for introducing me to this field.

I would like to take this opportunity to thank Dr Peter Rhodes, Dr Changjun Li, Dr Guihua Cui, Dr Li-Chen Ou, Dr Ji Wei, Dr Vien Cheung and my colleagues at the Department of Colour Science for their help, advice and more importantly for their friendship.

I would like to express my profound gratitude to all my friends and my entire family.

Abstract

Quantifying the appearance of coating products is essential in the automobile and automobile finishing industries for efficient product development and product/quality control. There is a specific need to develop techniques to measure the total appearance of metallic coatings. The present study focuses on two key attributes of visual texture: *coarseness* and *glint*. In order to develop models to capable of measuring these attributes, it was first necessary to design psychophysical experiments for assessing coarseness and glint as perceived on metallic-coating panels. The change in the appearance of the metallic coatings is known as a *gonioapparent effect*, and is greatly dependant on the illumination and viewing conditions. Therefore, appropriate conditions were carefully examined for the independent observation of coarseness and glint in order to discern those attributes. It was found that diffuse illumination was appropriate for viewing coarseness and directional illumination was appropriate for observing glint. Under these appropriately-controlled conditions, the perceptual coarseness and glint of sets of metallic-coating panels were assessed by human observers.

A digital camera was used to capture information on the spatial detail of the metallic-coating panels. An image of each panel was captured under the same viewing conditions as used for the visual assessments. The information in a single image was sufficient to represent a metallic-coating panel under identical diffuse illumination conditions for which observers assessed coarseness. For capturing information on glint, however, a high dynamic-range (HDR) image was necessary because the dynamic range of the scene in which the glint was observed exceeded that of the camera used in this study. Two computational models were developed to predict perceptual coarseness and perceptual glint by extracting associated features from the captured images. The performance of these models was verified by comparing predictions made using them with the perceptual coarseness and glint scaled by observers.

For industrial use, the visualisation of products on computer displays would give various opportunities, for example, to develop and design products or coatings and also to communicate appearance information. A digital camera and a suitable display would enable this to be achieved, but the ability to reproduce the appearance of metallic-coating products on displays in a satisfactory manner was found to have significant

challenges. The coarseness model developed in the present study was able to represent perceptual coarseness based on the images captured by the digital camera. However, the resolution of the images was not high enough to resolve the individual aluminium flakes contained in the coatings, which contribute to the visual texture. Therefore, verification of the images was carried out for the coarseness attribute by comparing the coarseness perceived in the images displayed on a liquid-crystal display (LCD) with the metallic-coating panels themselves. In addition to camera limitations, LCD resolution also prevented the same conditions used for physical panel assessment from being replicated. Therefore, two optimal conditions were selected and perceptual coarseness was scaled using images. Besides the difference in experimental conditions, there was also a difference in the “absolute” texture appearance between the two media because of errors in image reproduction of the images. In spite of this, the relatively-scaled perceptual coarseness for the image samples agreed well with that for the original physical samples. This implies that it is practicable to assess perceptual coarseness from an image on a display that simulates a metallic-coating panel.

Contents

Acknowledgements	i
Abstract	ii
Contents	iv
List of Figures	ix
List of Tables	xvi
Chapter 1 Introduction	1
1.1 Background.....	2
1.2 Appearance Attributes and Measurements.....	3
1.3 Aim of the study.....	4
1.4 Achievements.....	5
1.5 Thesis Structure	6
1.6 Publications.....	6
Chapter 2 Literature Survey	8
2.1 Introduction.....	9
2.2 Human Visual System.....	10
2.2.1 The Eye	10
2.2.2 Photoreceptors.....	12
2.2.2.1 Cone Fundamentals	15
2.2.2.2 Physiologically-Based Chromaticity Diagram	16
2.2.3 Spatial Properties of the Human Visual System.....	17
2.2.3.1 Luminance CSF	19
2.2.3.2 Luminance CSF with Chromatic Backgrounds	21
2.2.3.3 Chrominance CSF	22
2.3 Colorimetry	23
2.3.1 CIE 1931 Standard Colorimetric Observer	25
2.3.2 CIE 1964 Standard Colorimetric Observer	26
2.3.3 CIE XYZ Tristimulus Values.....	26
2.3.4 CIE xy Chromaticity Diagram	27
2.3.5 CIE 1976 UCS Diagram.....	28
2.3.6 CIELAB Colour Space and Colour Differences	29
2.3.7 S-CIELAB Colour Difference.....	30
2.4 Light-Field Formation and Measurements	31
2.4.1 Light Behaviour	31

2.4.2	Light/Colour Measurements	35
2.4.2.1	Spectroradiometer	36
2.4.2.2	Spectrophotometer	37
2.5	Metallic Coatings.....	37
2.5.1	Appearance of Metallic Coatings	40
2.5.1.1	Colour of Metallic Coatings.....	40
2.5.1.2	Visual Texture of Metallic Coatings.....	43
2.6	Texture Analysis.....	46
2.6.1	Statistic Methods	48
2.6.1.1	Spatial Grey Level Dependence Method	49
2.6.1.2	Grey Level Difference Method	51
2.6.1.3	Neighbouring Grey Level Dependence Method	52
2.6.2	Signal Processing Methods.....	53
2.6.2.1	Fourier Domain Filtering	54
2.6.3	Structural Methods	55
2.6.3.1	Grey Level Run Length Method	55
2.7	High Dynamic-Range Imaging.....	57
2.8	Image Acquisition and Display Devices.....	58
2.8.1	CCD Camera Systems	58
2.8.2	Display Systems	60
2.8.3	Device Characterisation.....	62
2.8.3.1	Linear and Polynomial Regression Model.....	65
2.8.3.2	GOG Model	67
2.8.3.3	Day, Taplin and Berns Method	69
2.9	Psychophysics: Quantitative Methods for Perceptual Responses.....	70
2.9.1	Categorical Judgement Scaling Method	71
2.9.2	Magnitude Estimation.....	74
2.10	Statistical Measures	75
2.10.1	Correlation Coefficient & Coefficient of Determination	76
2.10.2	Root Mean Square Error (RMSE)	77
2.10.3	Coefficient of Variation (CV).....	77
2.10.4	Confidence Intervals.....	78
2.10.5	The Wilcoxon Matched-Pairs Signed-Rank Test.....	78
2.11	Summary	80
Chapter 3	Assessing and Modelling Coarseness using Physical Samples.....	81
3.1	Introduction	82
3.2	Colour Measurement Instruments.....	83
3.2.1	Minolta CS1000 Tele-Spectroradiometer.....	83

3.2.2	GretagMacbeth ColorEye 7000A Spectrophotometer.....	84
3.3	Visual Assessments of Coarseness on Metallic-Coating Panels	85
3.3.1	Samples	86
3.3.2	Viewing Conditions	86
3.3.2.1	Experimental Settings.....	87
3.3.3	Scaling Perceptual Coarseness	90
3.3.4	Data Analysis and Results	91
3.3.4.1	Scale Value.....	91
3.3.4.2	Observer Variability	95
3.3.5	Discussion	97
3.4	Computational Model for Perceptual Coarseness Prediction.....	99
3.4.1	Capture and Measurement of the Target	102
3.4.2	Colour Space Transformation	104
3.4.2.1	XYZ Colour Space	104
3.4.2.2	LMS Colour Space	110
3.4.3	Contrast Measure	111
3.4.4	Applying CSF	112
3.4.5	Normalisation.....	113
3.4.6	Non-Linearity.....	115
3.5	Model Performance.....	116
3.6	Summary	119
Chapter 4	Assessing and Modelling Coarseness using a Display.....	122
4.1	Introduction.....	123
4.2	Display Evaluation	124
4.2.1	Temporal Suitability.....	125
4.2.2	Additivity Test	126
4.2.3	Chromaticity Constancy of Primaries	127
4.2.4	Spatial Independency	128
4.2.5	Spatial Uniformity.....	129
4.3	Display Characterisation	130
4.3.1	Data Sets	131
4.3.2	Implementation of Characterisation Models	131
4.3.2.1	GOG Model.....	131
4.3.2.2	Linear and Polynomial Model	132
4.3.2.3	Day, Taplin and Berns Model.....	133
4.3.3	Display Characterisation Performance	133
4.4	Image Reproduction.....	137
4.5	Visual Assessment using Image Samples	142

4.5.1	Comparison of Coarseness Differences	144
4.5.2	Scaling Perceptual Coarseness in Condition 6.....	147
4.5.3	Scaling Perceptual Coarseness in Condition 1.....	147
4.6	Data Analysis and Result.....	147
4.6.1	Observer Variability	148
4.6.2	Comparisons of Perceptual Coarseness Scaled in Different Conditions.....	152
4.6.3	Comparisons of Perceptual Coarseness and Coarseness Model Predictions	157
4.7	Summary	160
Chapter 5 Assessing and Modelling Glint using Physical Samples.....		163
5.1	Introduction	164
5.2	Visual Assessments of Glint.....	165
5.2.1	Samples.....	165
5.2.2	Viewing Conditions	166
5.2.2.1	Experimental Settings	166
5.2.3	Scaling the Perceptual Glint	168
5.2.4	Data Analysis and Results	170
5.2.4.1	Scale Value	170
5.2.4.2	Observer Variability.....	171
5.2.4.3	Glint Angle.....	179
5.2.5	Discussion.....	181
5.3	Computational Model for Perceptual Glint Prediction	184
5.3.1	Measurement and Capture of Targets	185
5.3.2	HDR Image Capture	186
5.3.2.1	LCR Image Capture	189
5.3.2.2	Linearisation of Camera Response.....	190
5.3.2.3	Spatial Uniformity Correction.....	193
5.3.2.4	Deriving Camera Response Function.....	193
5.3.3	Colour Space Transformation.....	195
5.3.4	Glint Feature Extraction	200
5.3.4.1	Contrast Measure	200
5.3.4.2	Segmentation.....	201
5.3.4.2.1	Iterative Method.....	203
5.3.4.2.2	Distribution-Estimate Method	204
5.3.4.3	Glint Features – Statistical Approaches	204
5.3.4.4	Non-Linearity.....	206
5.4	Model Performance	207
5.4.1	Model Performance: Iterative Method.....	208
5.4.1.1	Model Performance: Distribution-Estimate Method.....	214
5.5	Summary	220

Chapter 6	Conclusions	222
6.1	Overview.....	223
6.1.1	Assessing Coarseness using Physical Samples	223
6.1.2	Computational Model for Coarseness Prediction.....	224
6.1.3	Assessing Coarseness using a Display	226
6.1.4	Assessing Glint using Physical Samples	227
6.1.5	Computational Model for Perceptual Glint Prediction.....	227
6.2	Discussion and Future Work.....	230
References.....	232
Appendices	245
Appendix I:	Images of Metallic-Coating Panels used for Coarseness Assessments	246
Appendix II:	Scale Values and Model Predictions from Coarseness Assessments	247
Appendix III:	Images of Metallic-Coating Panels used for Glint Assessments	251
Appendix IV:	Scale Values and Model Predictions from Glint Assessments	252

List of Figures

Figure 1-1: An example of coarseness appearance. The appearance of coarseness is dependant on a viewing distance.	4
Figure 2-1: (a) A schematic drawing of a horizontal cross-section of the right human eye (adapted from (Matlin & Foley, 2007)). (b) Structure of the retina (adapted from (Miller, 2007)).	10
Figure 2-2: The proportion of photons transmitted by various components of the eye as a function of wavelength. The data were based on measurements of freshly enucleated eyes. Each curve is the transmittance at the rear surface of the labelled structure, therefore showing the cumulative effects of all the layers up to that point. The data was collected by Boettner and Wolter (1962) and adapted from (Packer & Williams, 2003).	12
Figure 2-3: The CIE 1924 luminance efficiency $V(\lambda)$ (solid line) and the CIE 1951 luminance efficiency $V'(\lambda)$ (dashed line) (CVRL, 2007).....	14
Figure 2-4: Densities of rods and cones as a function of eccentricity from the fovea along a horizontal meridian of the human retina (Osterberg, 1935).....	15
Figure 2-5: Cone Fundamentals (Stockman & Sharpe, 2000; Stockman and Sharpe et al., 1999).	16
Figure 2-6: MacLeod-Boynton chromaticity diagram with locus of monochromatic stimuli.	17
Figure 2-7: Schematic diagram to show the luminance CSF (Mullen, 1985).	19
Figure 2-8: The contrast sensitivity varies with average luminance levels (van Meeteren and Vos, 1972). Each curve shows a CSF at different average luminance levels from 10^{-4}cd/m^2 to 10cd/m^2 . At the low luminance levels (from 0.01 cd/m^2 to 0.0001 cd/m^2), the CSFs exhibit a lowpass filter. At the luminance levels from 0.1 cd/m^2 to 10 cd/m^2 , the CSFs exhibit a bandpass filter.....	20
Figure 2-9: The contrast sensitivity varies with angular display sizes (Carlson, 1982). Each curve shows a CSF at different angular display sizes. The contrast sensitivity increases with the increase of an angular display size.	21
Figure 2-10: Computational model prediction for mean luminance of 50 cd/m^2 with various saturations (Westland, 2005).	22
Figure 2-11: Schematic diagram of the chrominance CSF: red-green (dashed line) and yellow-blue (dotted line) together with the luminance CSF (solid line) (Mullen, 1985).....	23
Figure 2-12: An example of the viewing field of a typical colour matching experiment.....	23
Figure 2-13: Colour matching functions; $\bar{r}(\lambda)$, $\bar{g}(\lambda)$ and $\bar{b}(\lambda)$ (CVRL, 2007).....	25
Figure 2-14: Colour matching functions; $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$: the CIE 1931 Standard Colorimetric Observer (solid line) and the CIE 1964 Standard Colorimetric Observer (dashed line) (CIE, 2004b).....	25
Figure 2-15: xy chromaticity diagrams of the CIE Standard Colorimetric Observer. (a) The CIE 1931 Standard Colorimetric Observer (solid line) and the CIE 1964 Standard Colorimetric Observer (dashed line). (b) Chromaticity diagram with the position of the additive mixture of $\alpha A + \beta B$	28

Figure 2-16: A flowchart of S-CIELAB computational process (Zhang & Wandell, 1997).	30
Figure 2-17: Relationship between incident light, specular reflection and transmitted light.....	31
Figure 2-18: Microscopic view of light travelling through a layer.	33
Figure 2-19: Light interference. (a) Constructive interference. (b) Destructive interference.	34
Figure 2-20: Light reflection of three types of surface. (a) Mirror. (b) Semi-matt. (c) Matt.	34
Figure 2-21: Schematic diagrams showing the CIE recommended illumination and viewing geometries: left 0/45 and 45/0, right d/8 and 8/d (Luo, 2002).....	36
Figure 2-22: Coating layers on an automotive panel.	38
Figure 2-23: Light behaviour caused by the orientations of aluminium flakes. (a) Parallel alignment of the aluminium flakes. (b) Random alignment of the aluminium flakes.	39
Figure 2-24: Relationship between aluminium flake size and light reflection.....	39
Figure 2-25: Classification of five images contained different characteristics of texture.....	47
Figure 2-26: (a) An image consisting of four distinct texture regions. (b) The four identified texture regions in the image - the boundary map.	47
Figure 2-27: (a) An image of the Brodatz textures. (b) The synthesized image based on parameters extracted from the image shown in (a) (Portilla & Simoncelli, 2000).	48
Figure 2-28: Indications of a directional parameter θ for 0° , 45° , 90° and 135° and examples of a pair with a pixel X.....	49
Figure 2-29: (a) A 4×4 image with four grey levels 0-3. (b) General form of and grey level spatial dependence matrix for an image with grey levels 0-3 – the elements stand for the grey levels x and y for two paired pixels. (c)–(f) Matrices $P_{d\theta}(x, y)$ with a distance parameter $d = 1$ for $\theta = 0^\circ$, 45° , 90° and 135° respectively.....	50
Figure 2-30: An example of a NGLDM. (a) A 5×5 image with four grey levels 0-3. (b) A matrix $P_{ad}(x, y)$ computed from the image in (a).....	52
Figure 2-31: The u, v dimensions of the Fourier transform may be split into concentric rings (left) and wedges (right) to encode spatial frequency and orientation information respectively.	55
Figure 2-32: An example of the Run Length Method. (a) A 4×4 image with four grey levels 0-3. (b) Run length matrices $P_\theta(x, y)$ for each direction $\theta = 0^\circ$, 45° , 90° and 135° respectively.	56
Figure 2-33: An HDR image (right) created from four LDR images (left) (CAVE, 2007).	58
Figure 2-34: Schematic diagram of Bayer pattern and its process.....	60
Figure 2-35: Schematic diagram of the physical structure of a typical LCD (TN liquid crystal and active matrix) (adapted from (LG Philips, 2007)).....	62
Figure 2-36: An example of device-dependent values. Different cameras give different RGB values from the same scene.	63
Figure 2-37: A worked flowchart of the Day, Taplin and Berns characterisation model (Day and Taplin et al., 2004).	70
Figure 2-38: Illustration of judgement distributions of stimulus and category boundaries.....	73
Figure 2-39: Examples of the relationships between data sets and the correlation coefficient or the coefficient of determination.	77
Figure 3-1: A flowchart for the study of the present chapter.	83
Figure 3-2: A typical viewing cabinet used for visual assessment.	87

Figure 3-3: (a) A schematic diagram of a DigiEye® Viewing Cabinet. The samples are placed on the base. Two light sources (D65 simulators) are positioned at two bottom corners at each side and both emit light to the walls. The two top corners have curved surfaces to reflect light uniformly onto the samples. An observer looked down onto the sample from the viewing window. (b) The arrangement of the samples as presented to the observers.	88
Figure 3-4: SPD of the light source specified by measuring SPD of a white ceramic standard tile using a Minolta CS1000 tele-spectroradiometer (TSR).....	88
Figure 3-5: Spatial uniformity evaluation of the illumination. (a) Locations of measurements in uniformity evaluation. (b) Luminance in cd/m^2 at each measurement point. (c) CIELAB ΔE^*_{ab} colour differences between the centre and each point.	89
Figure 3-6: Comparison of the scale values derived from the mean-category value method and the categorical-judgement method for all samples.	95
Figure 3-7: The viewing condition used for the coarseness assessment at Akzo Nobel (Kirchner and Kieboom et al., 2007).	98
Figure 3-8: Mean of the observers' z-scores of coarseness with 95 % confidence intervals (the more negative z-scores indicate the perceptually coarser texture).....	100
Figure 3-9: Changes of R^2 values for contrast from the spatial grey level dependence method when the distance parameter changes from 1 to 11.....	101
Figure 3-10: A flowchart of the main framework for the coarseness model.....	102
Figure 3-11: An illustration of the experimental condition for the measurement of the spectral properties using the TSR and the image capture using the digital camera.	104
Figure 3-12: Colour distributions of the 156 metallic-coating samples, 124 solid-colour-coating panels and 166 patches in the ColorChecker DC. (a) A CIELAB L^*C^* diagram. (b) A CIELAB a^*b^* diagram.....	106
Figure 3-13: An original image of a metallic-coating panel (top left) and its Fourier spectrum images for luminance (top right) and chromatic channels: red-green channel (bottom left) and yellow-blue channel (bottom right). Note that the DC component is in the centre of each Fourier spectrum image and that the spatial frequency increases from the centre to outwards.....	112
Figure 3-14: Perceptual coarseness values obtained from the visual assessment of grey samples (top row). Original images of the samples (middle row). Fourier spectrum images of the luminance channel (bottom row).....	114
Figure 3-15: Perceptual coarseness values obtained from the visual assessment of green samples (top row). Original images of the samples (middle row). Fourier spectrum images of the luminance channel (bottom row).....	114
Figure 3-16: An example of the relationship between Fourier energy and the lightness of images in the green colour samples which are the same samples as in Figure 3-15 having similar perceptual coarseness (a scale value of 7.70 – 7.80) of the perceptual coarseness obtained from the visual assessment.	115
Figure 3-17: Comparison between the output values before linearisation and the scale values of the perceptual coarseness values.	116
Figure 3-18: Comparison of the model predictions with the scale values of the perceptual coarseness for all the 155 metallic-coating panels.....	118

Figure 3-19: Comparison of the model predictions with the scale values of the perceptual coarseness for each colour metallic-coating panels: (a) grey colour samples, (b) blue colour samples, (c) green colour samples and (d) purple colour samples.	118
Figure 3-20: Comparison of the model predictions with the scale values of the perceptual coarseness for all the 155 metallic-coating panels with 95 % confidence intervals.....	119
Figure 4-1: (a) SPD of the primaries of the LCD used in this experiment. (b) Chromaticity coordinates of the primaries of the LCD (solid line) and sRGB (dashed line) in the CIE 1976 Uniform Chromaticity Scale diagram.	125
Figure 4-2: Temporal suitability of the display expressed in CIELAB ΔE^*_{ab} from the last measurement.	126
Figure 4-3: Additivity errors in terms of ΔE^*_{ab} , ΔL^* , Δa^* and Δb^* : For example, ΔL^* = (the separate measurements of the mixture of the three primaries) – (the addition of the measurements the three primaries).	127
Figure 4-4: Chromaticity changes of each channel in CIE 1976 Uniform Chromaticity Scales diagram; (a) before; and (b) after black correction.....	128
Figure 4-5: (a) Values of CIELAB ΔE^*_{ab} and (b) CIEALB a^*b^* parameters of the grey patch displayed at the centre surrounded by the same grey background (×) and by each of white, black, red, green and blue backgrounds (filled circles).....	129
Figure 4-6: CIELAB ΔE^*_{ab} colour differences between the centre and nine locations evaluated for (a) white and (b) grey.....	130
Figure 4-7: Percentage CIELAB L^* lightness differences between the centre and nine locations evaluated for (a) white and (b) grey.	130
Figure 4-8: The radiometric scalars plotted against the digital counts for the red, green and blue channels. (a) The scalars from the measurements (×) and estimated values (full line) using the parameters from the GOG model. (b) The scalars from the measurements (×) and estimated values (full line) using the optimised transformation matrix from the Day, Taplin and Berns models.	135
Figure 4-9: A flowchart for image reproduction process.....	137
Figure 4-10: CIELAB colour difference between data sets MP, EC, EL and MI. (a) Comparisons of MP with EC, EL and MI. (b) Comparisons of EC with MP, EL and MI. (c) Comparisons of EL with MP, EC and MI. (b) Comparisons of MI with MP, EC and EL.	141
Figure 4-11: CIELAB (a) L^* (b) C^* and (c) a^*b^* differences between MP and MI.	142
Figure 4-12: Images used to test the errors of the output colours. The colour of the three images looks different, although they have the same ratio of black to white pixels.....	142
Figure 4-13: The viewing geometry used for comparison of the physical samples with the image samples (top) and the examples of the image sample on the LCD and parameters for Condition 1 and 6 (bottom).	146
Figure 4-14: Geometric mean perceptual coarseness difference for each of nine conditions together with the error bars. Coarseness difference five indicates that a pair of image samples displayed on the display had an equivalent coarseness difference to the physical sample pair presented in the viewing cabinet.	146

Figure 4-15: Comparisons between the scale values in Condition 6 using the image samples (labelled Condition 6), in Condition 1 (labelled Condition 1) and in the viewing cabinet using the physical samples (labelled Physical Sample).	155
Figure 4-16: Comparisons of the scale values of the samples in each colour group between Condition 6 and Condition 1.	155
Figure 4-17: Comparisons of the scale values of the samples in each colour group between Condition 6 and the physical samples.	156
Figure 4-18: Comparisons of the scale values of the samples in each colour group between Condition 1 and the physical samples.	156
Figure 4-19: The scale values (a) in Condition 6 using the image samples, (b) in Condition 1 using the image samples (c) in the viewing cabinet using the physical samples, plotted as a function of the associated model predictions.	159
Figure 4-20: Scale values for samples in individual colour groups in (a) Condition 6, (b) Condition 1, and (c) the viewing cabinet using physical samples plotted against their respective model predictions. Black dashed lines are linear regression lines (the scale values against the model predictions) for all samples in each condition. Red lines indicate linear regression lines for the actual plotted data.	159
Figure 5-1(a) The samples' arrangements presented to observers. (b) Experimental settings.	167
Figure 5-2: (a) Illumination and viewing geometry for the measurement. (b) SPD of the light source. .	168
Figure 5-3: Spatial uniformity evaluation of the illumination. (a) Geometry of measurements in uniformity evaluation. (b) Locations of measurements in uniformity evaluation. (c) Illuminance variation in lux at the 15 points.	168
Figure 5-4: Comparison of the scale values derived from geometric mean and arithmetic mean of raw observer data for all samples.	171
Figure 5-5: The results of the selected angles. (a) The indication of the angle measurement. (b) The maximum, minimum and mean angles selected at the each session (e.g., the result of Observer 1's first session is labelled as Ob1-1st). (c) 95 % confidence intervals of each session were plotted with their mean.	180
Figure 5-6: Sketch of the experimental set-up used at Akzo Nobel (Kirchner and Kieboom et al., 2007). An example of the angle of 45°.	182
Figure 5-7: A flowchart of the main framework for the glint model.....	185
Figure 5-8: An illustration of the experimental condition for the spectral properties measurements using a TSR and the image capturing using a digital camera.	186
Figure 5-9: (a) Response of camera at Exposure 1. (b) Response of camera at Exposure 2. (c) Illustration of regions A (Exposure 1 valid only), B (both exposures valid) and C (Exposure 2 valid only). (d) A combined HDR camera response.	188
Figure 5-10: Examples of LDR images captured with two exposure setting. (a) Images capture information of a dark part of the scene (the right hand side image). (b) Images capture a bright part of the scene (the left hand side images).	188
Figure 5-11: An example of selections of camera settings. If there are any patches are not covered with any of exposure settings, it results in missing parts of scene information (Exposure	

Setting Selection 1). On the other hand, there is no missing scene information, but only a limited range of information is included (Exposure Setting Selection 2).....	189
Figure 5-12: The camera responses (normalised in the range 0-1) of the green channel from two of the selected exposure settings (1/10 is shown in green and 1/400 is shown in blue) against the sum of the measured SPD of the greyscale patches. The plots circled with red dashed lines indicate the patches within the valid camera responses.	190
Figure 5-13: Camera responses of the valid patches of the greyscale for each red, green and blue channel plotted against the sum of the SPD of the corresponding patches (circles) with the fitted lines (solid lines) (graphs on the top row), and linearised camera responses converted using the power function (circles) with fitted lines (solid lines) (graphs on the bottom row) for the camera setting at the exposure time of 1/10 (Exposure 1).....	192
Figure 5-14: Camera responses of the valid patches of the greyscale for each red, green and blue channel plotted against the sum of the SPD of the corresponding patches (circles) with the fitted lines (solid lines) (graphs on the top row), and linearised camera responses converted using the power function (circles) with fitted lines (solid lines) (graphs on the bottom row) for the camera setting at the exposure time of 1/400 (Exposure 2).....	192
Figure 5-15: (a, b) Linearised camera responses against the sum of the SPD of the greyscale patches (green channel) for two exposures with the indications of the upper and lower end of the valid values. (c) The indications of the regions; A (Exposure 1 (exposure 1/10) valid only); B (both exposures valid); and C (Exposure 1 (exposure 1/400) valid only.) (corresponding to Figure 5-9(c)). (d) Normalised sum of the SPD against the HDR values combined from the two camera response of the greyscale patches (corresponding to Figure 5-9 (d)).....	194
Figure 5-16: Colour distribution of the 106 metallic-coating samples and the 166 patches in the ColorChecker DC. (a) A CIELAB L*C* diagram. (b) A CIELAB a*b* diagram.....	196
Figure 5-17: Comparisons of the predicted Y values from the models of $M = 3 \times 10$ and $M = 3 \times 3$. (a) Median values for each of the 106 samples. (b) Maximum for each of the 106 samples. (c) Minimum for each of the 106 samples.	200
Figure 5-18: Range of maximum and minimum luminance values in each image together with the mode (filled circle) – the luminance values were obtained from the linear characterisation model. The horizontal black dashed line indicates the luminance value of the white patch for reference.....	201
Figure 5-19: A histogram of an image that can be partitioned by a single threshold t	202
Figure 5-20: Luminance channel's histograms of one of the images from each grey, blue, brown, green, red and yellow colour samples (the luminance value was obtained using the linear characterisation model).	203
Figure 5-21: Histograms of solid-colour-coating panels.	204
Figure 5-22: An illustration of the theory of the distribution-estimate method.	204
Figure 5-23: (a) The position of 8-neighboring pixels of a pixel at coordinates (i, j) (8 shadowed cells). (b) An Example of a labelled image. It indicates three connected components in an image.....	206
Figure 5-24: A flowchart of a summary of the model components and their evaluation processes.....	208

Figure 5-25: A threshold t_i with the image histograms of three samples (grey, blue, red colour samples). The iteration is terminated at the threshold t_k , when all the pixels of the red colour sample become below the threshold.	209
Figure 5-26: The prediction errors of the models associated with the metrics M1-M8 at each threshold. A series of six graphs on the left, middle and right columns shows the errors in terms of R, R^2 and RMSE respectively. A series of three graphs in each row indicates the each of the six linearisation functions. The vertical axis presents the error and the horizontal axis indicates threshold t value. The results are based on the luminance values transformed from the HDR values using the linear regression characterisation model.	210
Figure 5-27: The prediction errors of the models associated with the metrics M1-M8 at each threshold as well as Figure 5-26. The results are based on the luminance values transformed from the HDR values using the polynomial regression characterisation model.	211
Figure 5-28: Comparisons of the perceptual glint against with the model predictions at the threshold of 56 (based on the luminance values obtained from the linear characterisation model). (a, b, c and d) The predictions of the model with of M1, M2, M5 and M6 respectively (before linearisation). (e and f) The predictions of the model with of M5 and M6 respectively (after the linearisation using Power2).....	213
Figure 5-29: The minimum values in the images of the samples were plotted against the value at their mode (based on the luminance value obtained from the linear characterisation model).	215
Figure 5-30: The model predictions against the perceptual glint. (a) The performance of the model incorporated with M2 and Power2 based on the luminance value obtained from the linear characterisation model. (b) The performance of the model incorporated with M2 and Power1 based on the luminance value obtained from the polynomial characterisation model.....	218
Figure 5-31: Model Predictions of the model incorporated with M2 and Power1 based on the luminance value obtained from the polynomial characterisation model plotted against the scale values for the samples in individual colour groups. Black dashed lines are linear regression lines to data (the scale values against the model predictions) for all samples. Red lines are liner regression lines of the plotted data.	219
Figure 6-1: A flowchart of the main framework for the developed coarseness model.....	225
Figure 6-2: A flowchart of the main framework for the developed glint model.	229

List of Tables

Table 2-1: A work example for the Wilcoxon matched-pairs signed-rank test.	80
Table 3-1: Repeatability test of Minolta CS1000 tele-spectroradiometer.....	84
Table 3-2: Accuracy test of the CE 7000A in terms of ΔE^*_{ab} of 10 tiles between the measurement data using the CE7000A and the calibration data provided by the NPL.	85
Table 3-3: Repeatability test of the CE 7000A in terms of mean, maximum and minimum ΔE^*_{ab} values of 10 successive measurements.....	85
Table 3-4: 1-9 categories used for the visual assessment.	91
Table 3-5: A worked example of the categorical-judgement method. In this example, the number of the sample, $m = 3$, the category, $n = 7$ and the total observations, $N = 8$, by 4 observers (twice by each observer).	93
Table 3-6: Observer repeatability and the associated mean, median, maximum and minimum values of all the samples and the individual grey, blue, green and purple samples in terms of R^2 and CV.	96
Table 3-7: Observer accuracy for each session and the associated mean, median, maximum and minimum values of all the samples and the individual grey, blue, green and purple samples in terms of R^2 and CV.	96
Table 3-8: A list of image texture analysis methods implemented and features extracted in the coarseness assessment using the fabric samples.	99
Table 3-9: Size of the transfer matrix and augmented matrices compared in this study.....	105
Table 3-10: A summary of the data sets and the performance of the characterisation model which showed the smallest test error for each training data set.	108
Table 3-11: Model performances using the Data Set 1.....	108
Table 3-12: Model performances using the Data Set 2.....	109
Table 3-13: Model performances using the Data Set 3.....	109
Table 3-14: Model performances using the Data Set 4.....	109
Table 4-1: Performance of the GOG models.	134
Table 4-2: Performance of the linear and polynomial regression models.....	134
Table 4-3: Performance of the Day, Taplin and Berns models.....	134
Table 4-4: The viewing conditions for the visual assessments.	144
Table 4-5: P-values by the Wilcoxon signed-rank test to investigate the significance of the differences between the coarseness difference of two physical samples and those of two image samples in each Condition 1, 2, 3, 4, 5, 6, 7, 8 and 9.....	146
Table 4-6: Observer repeatability: mean, median, maximum and minimum values of all observer data for coarseness assessment using Condition 6 and Condition 1, and the physical samples (from Chapter 3: Table 3-6).	149
Table 4-7: Observer accuracy: mean, median, maximum and minimum values of all observer data for the coarseness assessment using Condition 6 and Condition 1, and the physical samples (from Chapter 3: Table 3-7).	150
Table 4-8: Observer repeatability for each observer using Condition 6.	150

Table 4-9: Observer repeatability for each observer using Condition 1.....	151
Table 4-10: Observer accuracy for each session using Condition 6.....	151
Table 4-11: Observer accuracy for each session using Condition 1.....	152
Table 4-12: Measures of model performance in terms of R^2 and CV with the scaling factor SF (see Section 2.10.3) and the values of the slope α and the intercept β for a linear regression line fitted to data (the scale values against the model predictions).....	160
Table 5-1: A summary of observer accuracy measures from all the samples and the samples in each grey, blue, brown, green, red and yellow sample group.....	173
Table 5-2: A summary of observer repeatability measures from all the samples and the samples in each grey, blue, brown, green, red and yellow sample group.....	173
Table 5-3: Observer accuracy measures for each session from all the samples.....	174
Table 5-4: Observer accuracy measures for each session from the grey colour samples.....	174
Table 5-5: Observer accuracy measures for each session from the blue colour samples.....	175
Table 5-6: Observer accuracy measures for each session from the brown colour samples.....	175
Table 5-7: Observer accuracy measures for each session from the green colour samples.....	176
Table 5-8: Observer accuracy measures for each session from the red colour samples.....	176
Table 5-9: Observer accuracy measures for each session from the yellow colour samples.....	177
Table 5-10: Observer repeatability measures of each observer from all the samples.....	177
Table 5-11: Observer repeatability measures of each observer from the grey colour samples.....	177
Table 5-12: Observer repeatability measures of each observer from the blue colour samples.....	178
Table 5-13: Observer repeatability measures of each observer from the brown samples.....	178
Table 5-14: Observer repeatability measures of each observer from the green colour samples.....	178
Table 5-15: Observer repeatability measures of each observer from the red colour samples.....	178
Table 5-16: Observer repeatability measures of each observer from the yellow colour samples.....	179
Table 5-17: The scaling factor α and the exponents β for the linearisation of the camera responses, and R^2 values as indications of how well the power functions fit to the greyscale patches.....	193
Table 5-18: Training and test data for characterisation.....	197
Table 5-19: Model performances using the Data Set 1.....	199
Table 5-20: Model performances using the Data Set 2.....	199
Table 5-21: Model performances using the Data Set 3.....	199
Table 5-22: Top six models in terms of their performance in R^2 and RMSE. The results are based on the luminance values transformed from the HDR values using the linear regression characterisation model.....	210
Table 5-23: Top six models in terms of their performance in R^2 and RMSE. The results are based on the luminance values transformed from the HDR values using the polynomial regression characterisation model.....	211
Table 5-24: R values between the perceptual glint and the model predictions based on the luminance value obtained from the linear characterisation model.....	216
Table 5-25: R^2 values between the perceptual glint and the model predictions based on the luminance value obtained from the linear characterisation model.....	216
Table 5-26: RMSE values between the perceptual glint and the model predictions based on the luminance value obtained from the linear characterisation model.....	216

Table 5-27: R values between the perceptual glint and the model predictions based on the luminance value obtained from the polynomial characterisation model.....217

Table 5-28: R² values between the perceptual glint and the model predictions based on the luminance value obtained from the polynomial characterisation model.217

Table 5-29: RMSE values between the perceptual glint and the model predictions based on the luminance value obtained from the polynomial characterisation model.217

Chapter 1

Introduction

1.1 Background

In the modern automobile industry, textured coatings for cars are essential for attracting customers. The appearance of these coating products involves a variety of properties of perceptual attributes (colour, gloss, texture, *etc.*) and therefore there is no satisfactory method to describe total appearance. Due to the complexity of appearance measurement, product quality control often requires on the instrumental measurements of a limited number of attributes and relies heavily on human inspectors. To improve efficiency, there is considerable interest in being able to make objective and quantitative measurements of the appearance of products. As well as the measurement of appearance, its visualisation for products presented on computer displays is becoming more and more demanding, not only for the efficiency of product development and product/quality control, but also for the communication of products in a global sense.

The measurement of the appearance of coatings can be achieved from an investigation of the relationships between the appearance attributes perceived by observers and the physical parameters obtained by instrumental measurements. The numerical expression of this relationship allows perceptual attributes to be predicted from physical measurements. The appearance attributes perceived on coatings can be quantified by visual assessment. On the other hand, there are various possibilities for measuring physical parameters; these are dependent on physical properties of the coatings and, specifically, their optical properties. Physical properties are related to the structure and components of coatings. Optical properties are the result of the interactions between light and the components of the coatings. In some cases, such as in the surface-refinishing industry, only the optical properties can be measured, since information of the physical properties is not available. The present study, which was carried out in conjunction with the National Physical Laboratory and Akzo Nobel Coatings and Car Refinishes, is focused on the modelling of correlations between perpetual appearance attributes and measurements of the optical properties in order to quantify the appearance of metallic coatings.

Visualisation of appearance can be accomplished by displaying an image or video captured by a suitable imaging system such as a digital camera or by synthesising the appearance of coatings. The simplest method is to display a captured image. Advances in systems for image acquisition and reproduction have led to enhancements in the realistic depiction of the appearance of products; however, detailed textural information

which can be captured and displayed is still unsatisfactory because of limitations with current technologies. In order to be able to use images in practice, it is necessary to investigate the performance of current imaging and displaying systems at visualising appearance.

1.2 Appearance Attributes and Measurements

Some of the important and possible measurements of optical properties are *colour*, *gloss* and *texture*, although these measurements are not necessarily independent (Pointer, 2003).

The study of colour in coatings has been active and colorimetry is well established as a method to quantify colour (see Section 2.3 and 2.5.1.1). Spectral information of light reflected from the surface can be associated as the physical parameter and it is possible to make measurements using commercial instruments such as a spectroradiometer or a spectrophotometer. Colorimetric parameters can be derived based on *colour matching functions* (see Section 2.3) so as to correlate with the quantity of perceptual colours. Moreover, a *colour appearance model* (CIE, 2004a) is capable of predicting colour-perceptual attributes, for example, lightness, chroma, and hue, by taking into account the viewing conditions.

Gloss perception is associated with the way the light reflects from the surface of an object at and near the specular direction (Pointer, 2003). A gloss meter or a goniophotometer is often used for this measurement, which gives a relative intensity of reflected light as a function of illumination and viewing angle. However, unlike the colour appearance model, a standardised model which is correlated with human perception of gloss has yet to be established.

Texture is a term that refers to the spatial properties representing the surface of an object. The human response to texture can generally be described in terms such as coarseness, fineness, graininess, smoothness, roughness, mottle, speckle, *etc.* In terms of human recognition, texture is perhaps intuitively obvious, but it has, as yet, no precise definition. Typically, a digital image can be a representation of a two-dimensional spatial distribution of a surface and this fact can be utilised as an instrumental method. Attributes of texture can then be characterised by analysing the spatial distribution of the pixels in the image using any one of a number of various

methods depending on the attributes of interest and also on the purpose of measurement. However, because of the wide variability and interaction of texture attributes, it is often difficult to tell what actual physical characteristics in an image contribute to a particular appearance attribute. Moreover, resolution dependence of the texture attributes also makes the measurements difficult, *i.e.*, the perceived texture depends on the viewing distance and similarly, the texture elements in an image are dependent on the resolution of the components of the imaging system used (Gibson, 1950). For instance, Figure 1-1 illustrates resolution dependence of the texture as an example of a textural attribute of *coarseness*. Although it is a same object, when it is viewed from a far distance, the texture looks very fine, while it is perceived to be much coarser at a closer distance.

As can be seen, the measurement of total appearance requires investigation to isolate the separate appearance attributes. It is not known how many of these attributes exist or need to be quantified to characterise total appearance. However, in the present study, two texture attributes, *coarseness* and *glint* of the metallic coatings have been investigated.

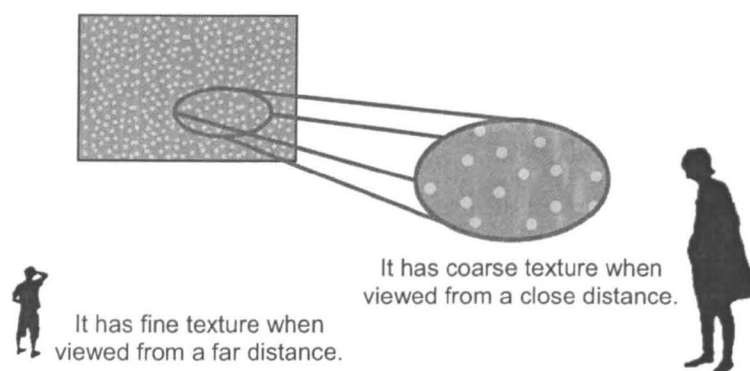


Figure 1-1: An example of coarseness appearance. The appearance of coarseness is dependant on a viewing distance.

1.3 Aim of the study

The aim of the present study was to model the appearance of attributes related to the texture of a number of sets of metallic coatings which were supplied by Akzo Nobel. The overall appearance of these metallic coatings consists of combinations of various attributes. In general, many studies have been carried out on measurements and visualisation or reproduction of the colour of metallic coatings (Alman, 1984; ASTM, 2003; McCamy, 1996; Venable, 1987; Westlund & Meyer, 2001) because of the large

impacts of gonioapparent colours (see Section 2.5.1.1). Compared with colour, there has been considerably less study on texture. Texture cannot be explained by a single parameter because it consists of a combination of various spatially related attributes. Although many texture attributes can be observed on metallic coatings *e.g.*, coarseness, glint, mottle, luster, the present study is focused on just two, *coarseness* and *glint*, which are identified as the key attributes by our collaborator, Akzo Nobel (2004-2006). The aims of this study are:

- To design an experiment to visually assess the coarseness and the glint of metallic-coating panels.
- To develop computational models capable of predicting the perceptual coarseness and the perceptual glint of the metallic-coating panels based on an image captured by a digital camera.
- To verify the information derived from the captured images in terms of the appearance of the coarseness.
- To investigate the possibility of using displayed images for the visual assessment of coarseness.

1.4 Achievements

All of the aims established at the beginning of the project (see Section 1.3) have been successfully achieved. They are summarised below.

- An experimental database for assessment of the coarseness and the glint of metallic-coating panels has been produced. This database includes data from observations made using a number of observers and samples, as well as images of those samples.
- Models for predicting the coarseness and glint from parameters derived from digital images were developed.
- Essential information necessary to design an imaging system for the measurement and display of metallic-coating panels was investigated.
- Limitations of this visualisation system were realised.

1.5 Thesis Structure

This thesis consists of six chapters. A brief account of each chapter is given below.

Chapter 1 covers the background, aims and overview of the present study.

Chapter 2 reviews the literature related to the present study. It includes an overview of human vision and colorimetry; background knowledge on metallic coatings and their appearance, texture analysis methods, image acquisition and display devices, as well as relevant psychophysics and statistical analysis tools.

Chapter 3 presents the investigations of the perceptual coarseness of the metallic-coating panels. The visual assessments to scale perceptual coarseness and the computational model developed for predicting perceptual coarseness are described.

Chapter 4 discusses the images used in Chapter 3 and also used to visualise the appearance of metallic coatings focusing on the coarseness attribute. The experiments that were carried out in order to achieve the aims, such as the reproduction of images of the metallic-coating panels and the visual assessments using those image displayed on the LCD are presented.

Chapter 5 introduces the work on the appearance of the glint on the metallic-coating panels. The visual assessments carried out in order to scale perceptual glint and the development of a computational model based on a generated HDR image are described.

Chapter 6 summarises the conclusions of the studies presented in chapters 3 to 5 and considers future work that could further our understanding of the appearance of gonioapparent coatings.

1.6 Publications

The following publications relate to study presented in this thesis.

1. Kitaguchi S, Westland S, Owens H, Luo MR & Pointer MR (2004). Surface Texture – A review. *National Physical Laboratory Report DQL-OR*, Teddington, Mdx., UK.

2. Kitaguchi S, Westland S & Luo MR (2005). Suitability of Texture Analysis Methods for Perceptual Texture. In *Proceedings of AIC Colour 2005, the 10th Congress of the International Colour Association*, Granada, Spain, 923-926.
3. Kitaguchi S, MacDonald L & Westland S (2006). Evaluating contrast sensitivity. In *Proceedings of SPIE Vol. 6057, Human Vision and Electronic Imaging XI*, San Jose, USA, 22-31.
4. Kitaguchi S, Luo MR, Kirchner EJJ & van den Kieboom GJ (2006). Computation model for perceptual coarseness prediction. In *Proceedings of Third European Conference on Colour in Graphics, Imaging, and Vision (CGIV)*, Leeds, UK, 278-282.
5. Kitaguchi S, Luo MR, Westland S, Kirchner EJJ & van den Kieboom GJ (2006). Assessing texture difference for metallic coating on different media. In *Proceedings of the 14th Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications*, Scottsdale, USA, 197-202.
6. Kitaguchi S, Westland S, Luo MR, Kirchner EJJ & van den Kieboom GJ (2008). Application of HDR colour imaging to modelling of glints in metallic coatings. In *Proceedings of AIC Colour 2008, Effects and Affects*, Stockholm, Sweden.

Chapter 2

Literature Survey

2.1 Introduction

This chapter provides background information related to the present study. Understanding the human visual system is fundamental for the study of appearance. Section 2.2 presents a brief review of the human visual system, including aspects of optical and neural processing, as well as the characteristics of human vision described on the basis of psychophysics. Section 2.3 covers colorimetry for specifying colour and colour differences. The appearance of visual texture originates in the interaction between light, a surface material and the human visual system. Therefore, it is also essential to understand how light is absorbed, transmitted or reflected by a material, and especially its surface, as reviewed in Section 2.4.1. The CIE recommended methods for the measurement of light, as well as suitable instruments are reviewed in Section 2.4.2. In Section 2.5, the behaviour of light is described regard to metallic-coating materials. As a result of the light reflected by a surface, colour and texture, *etc.* can be perceived. Metallic coatings are often called gonioapparent material since their appearance changes with the illumination and viewing geometry. This is much in the literature relating to this angular dependence and this is reviewed in Section 2.5.1.1. In comparison, the visual texture of metallic coatings has received less attention and the characteristics of visual texture are introduced in Section 2.5.1.2. Although there is a limited amount of literature available specifically on the texture of metallic coatings, texture in general has been studied and methods of texture analysis have been proposed, in order to extract texture-based features from digital images. Some of these methods are reviewed in Section 2.6 to demonstrate various approaches to the derivation of models for coarseness and glint prediction. In this study, digital images are utilised to extract information about visual texture. The dynamic range of conventional image-capture devices, for example digital cameras, is capable of capturing images of a typical scene. However, due to limitations in most of these devices, it is not possible to capture the full dynamic range of a scene containing very bright areas as, for example, caused by glint. Therefore, an HDR (high dynamic-range) image, as described in Section 2.7, is needed to be employed to obtain useful information about glint. In Section 2.8, a general introduction is provided on the digital camera and liquid-crystal display (LCD), used for the image acquisition and display in this study. Since these devices deal with the device-dependent values, a transformation into device-independent values is necessary in order to relate with colorimetric values. Characterisation methods, as describe in Section 2.8.3, are usually used for this purpose. The investigation of visual texture in

this study is based on psychophysics and a review of the methods employed is given in Section 2.9. Finally, the statistical measures for analysing the results obtained from the experiments are described in Section 2.10.

2.2 Human Visual System

2.2.1 The Eye

There are three basic components for the human perception: sources of light, objects and observers. The eye is the first place where reflected light from an object illuminated by a light source reaches. A brief description of the human visual system is provided in the following sections.

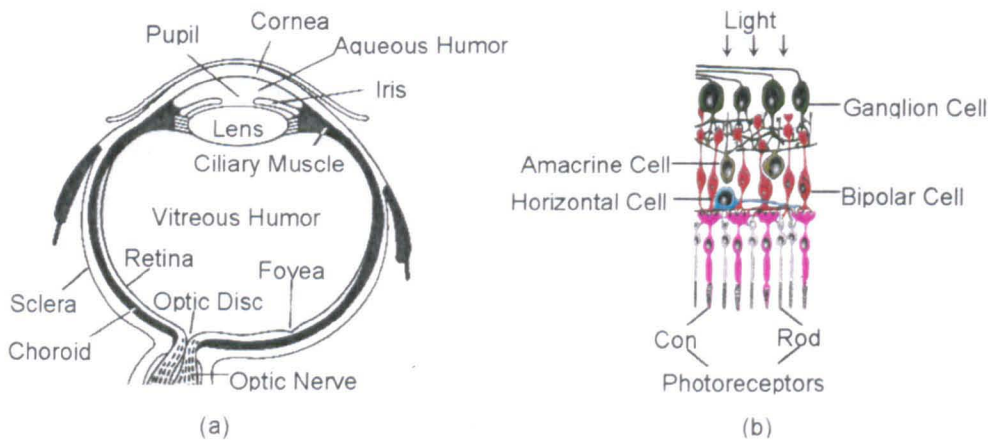


Figure 2-1: (a) A schematic drawing of a horizontal cross-section of the right human eye (adapted from (Matlin & Foley, 2007)). (b) Structure of the retina (adapted from (Miller, 2007)).

Figure 2-1 (a) shows a horizontal cross-section of the right human eye. The *cornea* is the front part of the eye where light first enters. It is approximately spherical with a radius of curvature of about 8 mm, its thickness is about 0.5 mm and its refractive index is approximately 1.38 (Atchison & Smith, 2000). The cornea is nearly transparent tissue without blood vessels and is contiguous with the *sclera*, the tough protecting envelop of the eye in which is inserted the extrinsic muscles that move the eye in its orbit (Wysecki & Stiles, 2002). The cornea absorbs about 10 % to 20 % of the incident light in the visible spectrum (approximately between 380 nm and 780 nm) but more than 99 % of the incident light in the ultra-violet region (<300 nm) (Packer & Williams, 2003). Figure 2-2 indicates the light loss due to the absorption of the cornea

and other components of the eye. The data is based on the measurements of freshly enucleated eyes (Boettner & Wolter, 1962). Between the cornea and the *lens* are the *aqueous humour* which is watery fluid with refractive index of about 1.33 (Atchison & Smith, 2000) and the *iris* which is an opaque muscular diaphragm forming a circular aperture, the *pupil*, through which the light passes. The size of the pupil is regulated by muscles in the iris and varies from 2 mm to 8 mm in diameter (Hunt, 1998). When bright light hits the eye, the ring-shaped sphincter muscle in the iris contracts (shrinks) the pupil. In dim light, the spindle cells dilate the pupil. The lens which is suspended between the aqueous humour and the *vitreous humour* (which is watery fluid with refractive index of about 1.33), is a biconvex structure made of thousands of roughly concentric layers with refractive index varying from 1.38 to 1.41 (Saxby, 2002). Together with the cornea as well as the other components of the eye, the lens refracts the light, and as a result helps the eye to focus the image-forming light at the *retina*. Among two main refracting components: the cornea and the lens, the cornea has the greater power. However, the corneal power is constant, whereas the power of the lens can be changed when the eye needs to focus at different distances. The process is called *accommodation* and occurs because of alteration in the lens shape by the *ciliary muscle* (Atchison & Smith, 2000). The amount of refraction that occurs is also a function of wavelength and therefore the visual system can only focus perfectly at a particular wavelength, which is about 580 nm in the green part of the visible spectrum. Therefore, light with shorter and longer wavelengths cannot be focused. This effect is known as *chromatic aberration* (Wysecki & Stiles, 2002). The pigments of the lens absorb light primarily at short wavelengths and absorb less than 10 % at wavelengths between 450 nm and 900 nm, in young adults (Packer & Williams, 2003) but these amounts vary between observers, and within the same observer with age (Malacara, 2002).

Transduction of light into neural signals is a chief function of the retina. The retina and the *optic nerve* originate as outgrowths of the developing brain. Hence, the retina is part of the central nervous system. The retina is thin layers of nerve tissue lining most of the *choroid*, vascular and pigmented layer attached to the sclera. The thickness of the retina varies from 0.05 mm at the *foveal centre* to about 0.6 mm near the *optic disc* (Atchison & Smith, 2000). An area where the optic disc is (about 3° across), is known as the *blind spot* because of the lack of *photoreceptors* to respond to the light, thus there is a break in the visual field. The centre area of the retina (the *fovea*) is covered by the *macula* which is a layer of the yellow pigment that is located

between the photoreceptors and *bipolar cells* (Figure 2-1 (b)). It is most intense in the fovea and gradually fading out beyond the fovea (Wyszecki & Stiles, 2002). It mostly absorbs light primarily at short wavelengths (Figure 2-2).

As shown in Figure 2-1 (b), the layers in the retina incorporate the neurons which can be classified into five classes: *ganglion cells*, *amacrine cells*, *bipolar cells*, *horizontal cells* and *photoreceptors*. The deepest layer of neurons, the photoreceptors, processes the light first (Lee, 2005). The photoreceptors convert light into neural signals and then transmit these signals to the bipolar cells and on to the ganglion cells. It is only the axons of these ganglion cells that are collected in a bundle at the optic disc, and leave the eye to form the optic nerve. In addition to this direct pathway from the photoreceptors to the brain, two other kinds of cells contribute to the processing of visual information in the retina. In this indirect pathway, the horizontal cells are interposed between the photoreceptors and the bipolar cells, and the amacrine cells between the bipolar cells and the ganglion cells.

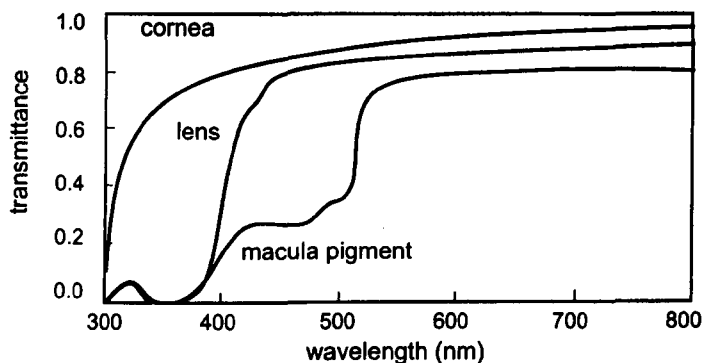


Figure 2-2: The proportion of photons transmitted by various components of the eye as a function of wavelength. The data were based on measurements of freshly enucleated eyes. Each curve is the transmittance at the rear surface of the labelled structure, therefore showing the cumulative effects of all the layers up to that point. The data was collected by Boettner and Wolter (1962) and adapted from (Packer & Williams, 2003).

2.2.2 Photoreceptors

There are two types of retinal photoreceptors, rods and cones (Figure 2-1 (b)). The rods and cones differ in a number of ways. The arrangement of the circuits that transmit the rod and cone information to the ganglion cells in the retina is different. The early stages of the pathways that link the rods and the cones to the ganglion cells are largely independent (Lee, 2005). The pathway from the rods to the ganglion cells

involves a distinct class of the bipolar cell which is called *rod bipolar cell*. Each rod bipolar cell contacts between 15 and 50 or more rods, depending on eccentricity. In contrast, the cone system is much less convergent. The pathway from the cones to the ganglion cells involves two different classes of the bipolar cell which are known as the *midget bipolar cell* and the *diffuse bipolar cell*. A midget bipolar cell makes contact with only one cone, while a diffuse bipolar cell makes contact with many (usually 5 to 10) cones. Convergence makes the rod system a better detector of light, because small signals from many rods are pooled to generate a large response in the rod bipolar cell. At the same time, this convergence reduces the spatial resolution of the rod system, since the source of a signal in a rod bipolar cell could have come from anywhere within a relatively large area of the retina (Purves and Augustine *et al.*, 2001). On the other hand, the cone system has less detection of light, but the system, especially involved the midget bipolar cells, preserves a spatial resolution as fine as the individual cone because of the less convergence of the cone information (Purves and Augustine *et al.*, 2001). Thus, the rods are responsible for vision in low luminance levels, while the cones are responsible in high luminance levels. Vision relying on rods is referred to as *scotopic vision* (below about 0.001 cd/m^2) and on cones is referred to as *photopic vision* (above about 3 cd/m^2), and vision, between scotopic and photopic vision, in which both rods and cones are active is referred to as *mesopic vision* (Valberg, 2005). The sensitivities of human vision under conditions of photopic and scotopic vision in the visible spectrum have been characterised and are called as the CIE 1924 luminance efficiency $V(\lambda)$ and as the CIE 1951 luminance efficiency $V'(\lambda)$ respectively as shown in Figure 2-3. The peak shifts from 555 nm for photopic vision to the lower wavelength 507 nm for scotopic vision. This is called the *Purkinje shift* (Lee, 2005). For instance, red colours appear brighter than blue colours at high luminance levels, but these same blue colours appear brighter than red colours at low luminance levels. The luminance efficiency $V(\lambda)$ for photopic vision is originated in a combination of spectral responses of the three types of cones which are referred to as *long-*, *medium-* and *short-wavelength-sensitive* (also called the *L-*, *M-*, and *S-*cones), according to the peaks of their relative spectral sensitivities. The luminance efficiency $V'(\lambda)$ for scotopic vision is based on spectral response of the rods.

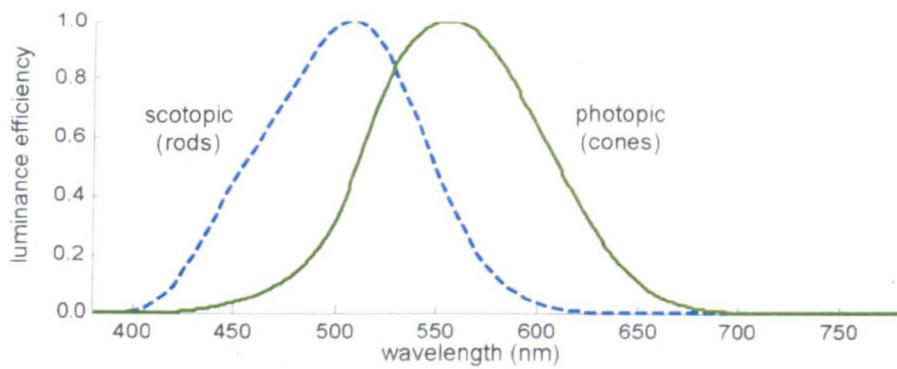


Figure 2-3: The CIE 1924 luminance efficiency $V(\lambda)$ (solid line) and the CIE 1951 luminance efficiency $V'(\lambda)$ (dashed line) (CVRL, 2007).

Another important difference between the rods and the cones is in their distribution across the retina (Wandell, 1995). As shown in Figure 2-4, the central fovea contains no rods, but it contains the highest concentration of the cones. Thus, the region of the highest spatial vision in the retina is the fovea. The presence of rods begins at an angle of about 4° from the central fovea. The distributions of the L -, M - and S -cones are also not the same. The S -cones are sparsely distributed in the retina and almost absent at the centre of the fovea (Curcio and Allen *et al.*, 1991). Additionally, the number of S -cones are considerably fewer than that of L - or M -cones (the $L:M:S$ cones are approximately in the ratio 40:20:1 (Walraven & Bouman, 1966)).

While the rods are incapable of providing colour vision, the cones are responsible for the generation of chromatic signals which lead to the perception of colour (Purves and Augustine *et al.*, 2001). However, individual cones are entirely color blind, like rods. Their response is simply a reflection of the number of photons of light that they capture, regardless of the wavelength of the photon. Colour vision is possible only when the photoreceptors of more than one work together by comparing the activity in different classes of cones. It is known that the three cone signals are compared and combined in the retina and, by the time the cone signals reach the ganglion cells, colour seems to be encoded in opponent-colour processes (Lee, 2005). Based on the responses of individual ganglion cells and also lateral geniculate nucleus (LGN) cells at higher levels in the visual pathway, three types of mechanisms have been found for the opponent-colour process: an achromatic mechanism and two chromatic mechanisms. The achromatic mechanism carries information on luminance (black and white) and responds information from M - and L -cones. One chromatic mechanism is responsible for red-green differences and processes information about differences between L - and

M-cones. The other chromatic mechanism is responsible for yellow-blue differences and processes the differences between *S*-cones and a combined signal from both *L*- and *M*-cones.

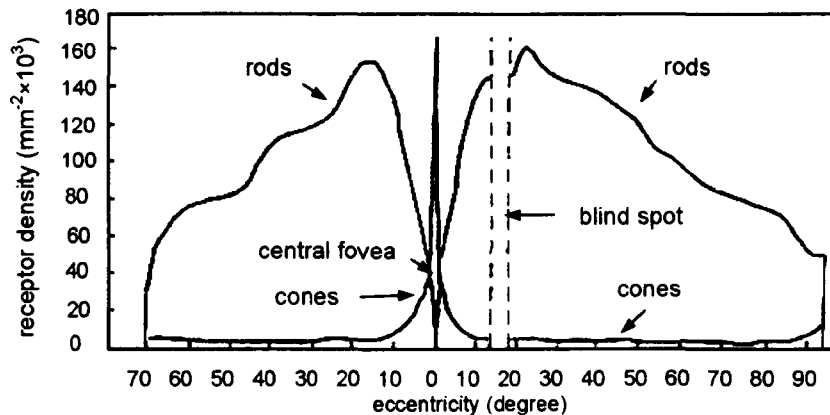


Figure 2-4: Densities of rods and cones as a function of eccentricity from the fovea along a horizontal meridian of the human retina (Osterberg, 1935).

2.2.2.1 Cone Fundamentals

Cone spectral sensitivity functions are called *cone fundamentals*. The cone fundamentals describe the sensitivity of the cones, but this is not the absorbance spectra of the cones because of the presence of other light absorbing elements, *e.g.*, the cornea, the lens, the macula, *etc* (Stockman & Sharpe, 1999). Various methods have been used to measure the cone fundamentals. Details of the experimental technique for the measurement of cone fundamentals are described by Stockman and Sharpe (1999). A set of cone fundamentals (Stockman & Sharpe, 2000; Stockman and Sharpe *et al.*, 1999) is shown in Figure 2-5. They are derived based on *L*- and *M*-cones spectral sensitivity measurements in single-gene red-green dichromats, or they are known protanopes (who are missing *L*-cone functions) and deuteranopes (who are missing *M*-cone functions), and *S*-cone spectral sensitivity measurements in *S*-cone monochromats and normal vision observers, and analysis of Stiles and Burch colour matching functions (Stiles & Burch, 1959).

Since the spectral sensitivities of each type of cone overlap, the methods for their measurement are not as simple and precise as the *colour matching methods* that will be described in Section 2.3. This is the reason that basic photometry and colorimetry were not derived in terms of cone sensitivities.

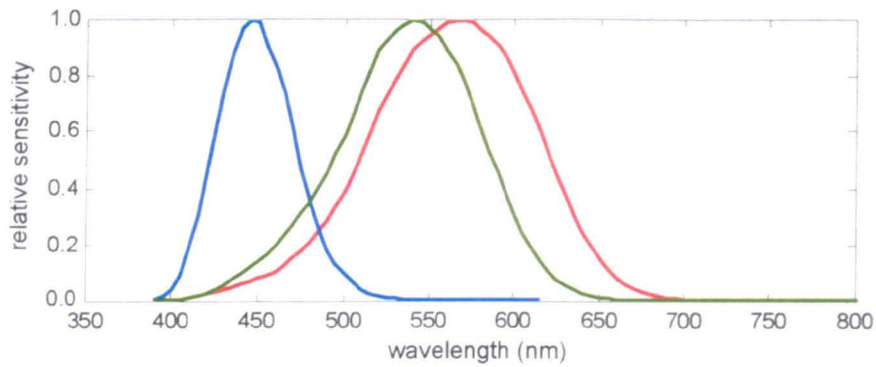


Figure 2-5: Cone Fundamentals (Stockman & Sharpe, 2000; Stockman and Sharpe et al., 1999).

2.2.2.2 Physiologically-Based Chromaticity Diagram

A physiologically-based chromaticity diagram was suggested by Maxwell (1860) who introduced a diagram comprised of an isosceles triangle with one cone fundamental at each corner. Although the Maxwell triangle has the advantage that it incorporates an axis corresponding to each cone type, it has the drawback that the axes are not orthogonal, each being perpendicular to a side of the triangle (MacLeod & Boynton, 1979).

MacLeod-Boynton (1979) suggested a chromaticity diagram which describes colours in terms of their relation with the *L*-*M*- and *S*-cones. This is a two-dimensional chromaticity diagram presented on a constant luminance plane. Assuming no contribution of the *S*-cone to the luminance, the luminance can be defined by the sum of the *L*- and *M*-cone sensitivities and chromaticity coordinates can be obtained by Equation 2-1.

$$\frac{L}{(L+M)}, \quad \frac{M}{(L+M)}, \quad \frac{S}{(L+M)} \quad \text{Equation 2-1}$$

The diagram, shown in Figure 2-6, consists of the relative excitation of the *L*-cones on the horizontal axis and the relative excitation of the *S*-cones on the vertical axis. This diagram has become one of the main methods for specifying colour stimuli in vision research, because of the advantage of the direct correlation of the coordinates to the cone fundamentals. Another advantage is that the central gravity rule for colour mixture can be applied in a straight-forward way; a mixture of two lights lies on the line between them at a distance proportional to the luminance of those lights (which are the coordinates of the diagram) (Benzschawel, 1992). However, its utility in terms of

colour specification has not been proved. A big disadvantage is that the diagram is perceptually highly non-uniform (MacLeod & Boynton, 1979).

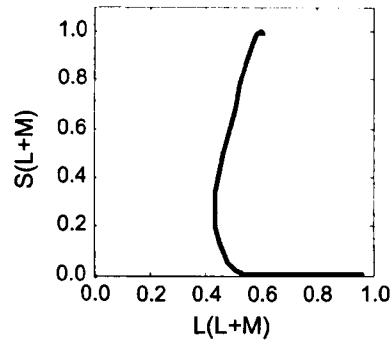


Figure 2-6: MacLeod-Boynton chromaticity diagram with locus of monochromatic stimuli.

2.2.3 Spatial Properties of the Human Visual System

The human visual system cannot only be considered in terms of spatially isolated stimuli, since objects are always seen in relation to their spatial configuration in the real-world. As has been shown in the previous sections, the human visual system processes information at many stages and each stage has various linear and non-linear characteristics. The fact that there are many things we do not understand about the neural processing in the eye/brain system leads to limitations in the description of human visual perception. Therefore, psychophysical measurements are often conducted in order to explore the various phenomena of visual perception. The spatial characteristics of the human visual system are typically explored through psychophysical measurements of *contrast sensitivity* which provide information about the ability of the eye to discriminate in a spatial sense, in terms of both luminance and chromaticity. The *contrast-sensitivity function* (CSF) is a measure of the contrast sensitivity of the human visual system as a function of a spatial frequency. The CSF resembles the *modulation transfer function* (MTF) associated with both in image science and optical systems. The MTF describes the amount of contrast reduction as a function of spatial frequency. If a sine-wave grating is passed through a perfect optical system, the image will still be a sine-wave grating but will have a slightly reduced contrast, or modulation, compared to the original object. If the eye only consisted of an optical system, the perceived luminance contrast of a sine-wave grating would be similar to the MTF. However, since the human visual system consists of both optical and neural processing elements, the characteristics of the CSF differ from that of the MTF. For instance, the luminance contrast attenuates at both high and low spatial

frequencies. One reason for the attenuation of contrast at low frequencies is the lateral inhibition within the retinal ganglion cells. A typical retinal ganglion cell presents a centre region with either excitation or inhibition and a surround region with the opposite sign. Therefore, when the spatial frequency is very low, a bright bar in a stimulus pattern covers both the inhibitory and the excitatory region of the ganglion cell. Moreover, it is known that the information passed through the retinal image can be discriminated according to its orientation and spatial/temporal frequency in further processes (such as in visual cortex, (Wandell, 1995)).

The contrast sensitivity of the human visual system is normally assessed by displaying sine-wave gratings of specific spatial frequencies and asking observers to adjust the contrast until they can just detect the presence of the grating against the uniform background (Westland, 2002). The usual definition of contrast of a sine-wave grating is given by the *Michelson contrast equation* (Michelson, 1927) as shown in Equation 2-2.

$$M = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \quad \text{Equation 2-2}$$

where M is a measure of contrast, I_{\max} and I_{\min} are the maximum and minimum intensities in the stimulus. Contrast sensitivity is then defined as the inverse of the contrast threshold.

It should be noted that contrast sensitivity is dependent on many factors including stimulus size, viewing distance, retinal position, age of an observer, eye movement and observation time. Therefore, the CSF is not unique. The extent to which the CSF is to account for vision in the everyday environment is unclear: in everyday vision, contrast levels are usually above threshold, but the CSF describes only visual thresholds and so one may question the usefulness of the function. However, it was found that the CSF incorporated with the metrics for image-difference and image-quality assessments seems to give reasonable results, although these are superthreshold applications (Barten, 1990; Bouzit & MacDonald, 2001; Sun & Fairchild, 2004).

In the following sections, the factors that affect the human contrast sensitivity and the CSF resulted in the influences of these factors are reviewed.

2.2.3.1 Luminance CSF

There have been many experiments involving the measurement of the luminance CSF (Campbell & Robson, 1968; Carlson, 1982; van Meeteren & Vos, 1972; Watanabe and Sakara *et al.*, 1976), but since the measurements were carried out using various experimental conditions, the data were not always comparable.

The typical shape of the luminance CSF for observers with normal vision, measured using stimuli consisting of a luminance-varying grating with central fixation at photopic levels can be seen in Figure 2-7 where the contrast sensitivity is plotted against the spatial frequency (cycles/degree). The highest sensitivity is usually found in the middle spatial frequency range at approximately 2–6 cycles/degree, with a sharp drop in sensitivity towards high spatial frequencies and reaching zero at about 60 cycles/degree (the point at which detail can no longer be resolved by the eye), and with a slower loss in sensitivity at low spatial frequencies; thus the CSF exhibits a bandpass filter. However, this shape applies only to the CSF measured using the variable luminance gratings. The CSF for chromatic stimuli often differs from this shape. Also, it is known that the sensitivity and shape of the functions change with various factors as listed above.

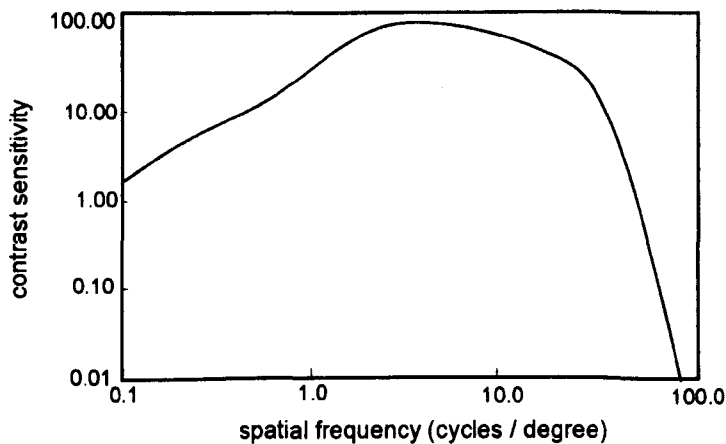


Figure 2-7: Schematic diagram to show the luminance CSF (Mullen, 1985).

The CSFs shown, in Figure 2-8, were measured by van Meeteren and Vos (1972) at various luminance levels from 10^{-4} cd/m² to 10 cd/m² using white light. As the luminance levels are progressively reduced, the contrast sensitivities generally reduce and the point of highest sensitivity moves to lower spatial frequencies. Finally, when the rods dominate vision, the shape becomes that of a lowpass rather than a bandpass filter (when less than 0.01 cd/m²). At mean luminance levels over 1 cd/m², the contrast

sensitivities at lower spatial frequencies remain constant (although Figure 2-8 shows the CSFs up to 10 cd/m^2 , this consistency at low luminance can be seen for the CSFs at more than 10 cd/m^2 (Wandell 1995)). The range in which contrast sensitivity becomes constant is called *Weber's law* regime (Wandell 1995). Weber's law can be expressed by $\Delta I/I = K$ where ΔI represents the difference in threshold, I is intensity of the background and K is a constant.

Weber's law for CSFs is imprecise because Weber's law does not hold at high spatial frequencies and at low mean luminance levels. However, it suggests the importance of contrast rather than absolute intensity (absolute luminance levels) for the human visual system. For example, the mean luminance levels in the experimental results shown in Figure 2-8 vary by a factor of 10^5 , yet the contrast sensitivity generally varies by only a factor of approximately 20. The pattern of results suggests that the visual system preserves the contrast sensitivity, as suggested by Weber's law, rather than absolute intensity.

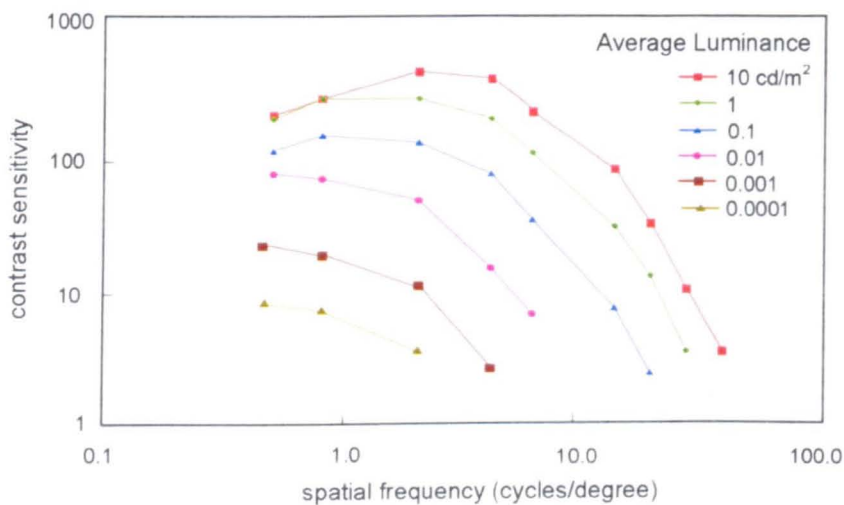


Figure 2-8: The contrast sensitivity varies with average luminance levels (van Meeteren and Vos, 1972). Each curve shows a CSF at different average luminance levels from 10^{-4} cd/m^2 to 10 cd/m^2 . At the low luminance levels (from 0.01 cd/m^2 to 0.0001 cd/m^2), the CSFs exhibit a lowpass filter. At the luminance levels from 0.1 cd/m^2 to 10 cd/m^2 , the CSFs exhibit a bandpass filter.

Another factor that affects the contrast sensitivity is the angular size. The contrast sensitivity increases as the angular display sizes become larger. Carlson (1982) investigated the importance of the angular display size as shown in Figure 2-9. The measurements were made for large ranges of angular display sizes extending from 0.5° to 60° . The luminance of the test object was 108 cd/m^2 with a surround luminance of one tenth of this value.

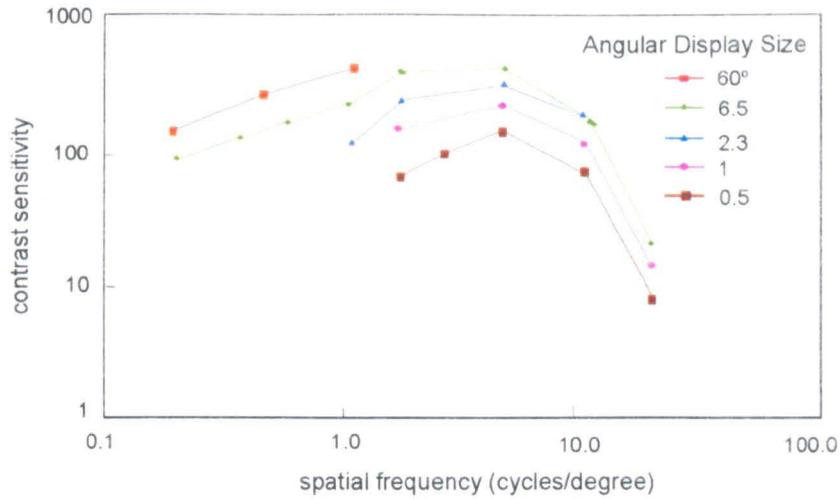


Figure 2-9: The contrast sensitivity varies with angular display sizes (Carlson, 1982). Each curve shows a CSF at different angular display sizes. The contrast sensitivity increases with the increase of an angular display size.

Besides these factors, the contrast sensitivities vary with the age of the observers (Iwata and Okajima *et al.*, 2001) and with retinal eccentricity (Virsu & Rovamo, 1979). The contrast sensitivities decrease for older observers, especially at higher spatial frequencies in comparison with younger observers. The previously introduced CSFs (in Figure 2-7 to Figure 2-9) are measured at the foveal vision. The contrast sensitivity tends to decrease at peripheral locations. One reason for these decreases in sensitivities is that the density of the cones falls off with visual eccentricity so that there are fewer sensors in the peripheral area to encode signals (Wandell 1995).

2.2.3.2 Luminance CSF with Chromatic Backgrounds

The contrast sensitivities of the luminance gratings measured using chromatic backgrounds are reported by Owens (2002). When the contrast sensitivity was measured using the fixed mean luminance of the gratings but the chromaticity of the field was varied from neutral to being chromatic in each of eight colour directions (red, yellow, green, blue, cyan, lime, purple and orange), it was evident that the sensitivity to luminance contrast was consistently less for chromatic stimuli than for achromatic stimuli as shown in Figure 2-10. This suggests that the human visual system does not completely process achromatic and chromatic information independently. In practice, achromatic and chromatic information does not usually appear separately. Several image-difference and image-quality metrics have been proposed to incorporate the

luminance CSF (Barten, 1990; Bouzit & MacDonald, 2001; Sun & Fairchild, 2004). These metrics may be overestimating the sensitivity of the visual system to luminance modulations when the metrics are applied to colour images.

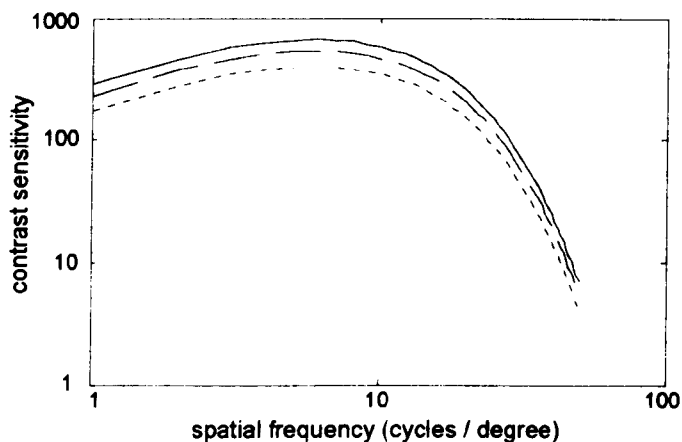


Figure 2-10: Computational model prediction for mean luminance of 50 cd/m^2 with various saturations (Westland, 2005).

2.2.3.3 Chrominance CSF

Typical chrominance CSF for isoluminant gratings (red-green and yellow-blue) is given in Figure 2-11 together with the luminance CSF. Different from the luminance CSF, the chrominance CSF exhibits a lowpass filter and significantly lower cut-off frequencies, and also the chrominance CSF for yellow-blue shows high-frequency attenuation at much lower spatial frequencies than is shown for the red-green due to the sparse retinal distribution of the *S*-cones (see Section 2.2.2). However, there are variations in the obtained results from the various experimental measurements. This may be caused by the fact that purely chromatic stimuli (isoluminant) gratings are not easy to produce. It should be noted that in colour imaging applications, it is impossible to separate image information into pure luminance and chrominance components. Therefore, for applications to image analysis, the precision of the chrominance CSF is not so critical (Lee, 2005).

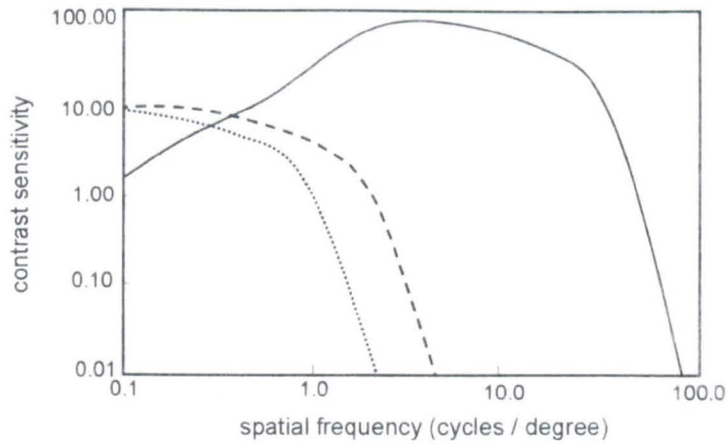


Figure 2-11: Schematic diagram of the chrominance CSF: red-green (dashed line) and yellow-blue (dotted line) together with the luminance CSF (solid line) (Mullen, 1985).

2.3 Colorimetry

Colorimetry is a scientific field concerned with the assignment of a concise numerical representation to any given colour stimulus. This is achieved by describing the spectral power distribution of any given colour stimulus by a much reduced number of parameters. Since three types of cones are used in normal human colour vision (see Section 2.2.2), it can be expected that three numbers are both necessary and sufficient. Therefore, colorimetry could aim to determine the three cone spectral response functions. However, in 1920s when this study was first attempted, it was not technically possible to measure the cone response functions so the three functions derived were not these functions but rather linear combinations of possible functions. The functions were derived by psychophysical experiments, a *colour matching experiment*, and are known as the *colour matching functions* (CMFs).

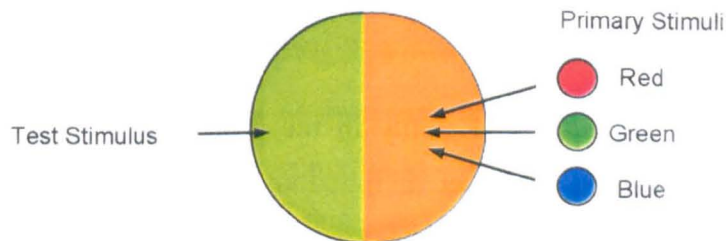


Figure 2-12: An example of the viewing field of a typical colour matching experiment.

One of the important results of this experiment was that almost all colours may be visually matched by a suitably-adjusted additive mixture of three different lights, termed *primary stimuli*. As given in Figure 2-12, the basic colour matching experiments use a bipartite viewing field (with a certain angular subtense), where three primary stimuli and a test stimulus are presented onto two separate halves. An observer adjusts the radiant powers of the primary stimuli until the mixture appears to match the test stimulus. When the radiant powers of the primary stimuli are measured for monochromatic test stimuli spanning the entire visible spectrum, the CMFs $\bar{r}(\lambda)$, $\bar{g}(\lambda)$ and $\bar{b}(\lambda)$ can be created, which are shown in Figure 2-13. Thus, a linear combination of the three primaries yields a colour stimulus $Q(\lambda)$ as follows.

$$Q(\lambda) = R\bar{r}(\lambda) + G\bar{g}(\lambda) + B\bar{b}(\lambda) \quad \text{Equation 2-3}$$

where R , G and B are scalars that represent the portions of the three fixed primaries that match a colour Q . R , G and B are called the *tristimulus values* of Q .

The choice of primary is any set such that no one primary stimulus can be matched by a mixture of the other two. The unit for the radiant power of the primaries is selected so that the mixture of one unit from each of the three primaries can be matched to the equal-energy white which is the stimulus having equal radiant power at every wavelength. For any real primaries, it is not always possible to match the test stimulus with combinations of the primaries; it is sometimes necessary to add one of the primaries to the test stimulus. This results in negative values for the CMFs. It is however, computationally simpler if no negative values are included. Thus, the International Commission on Illumination (*Commission Internationale de l'Éclairage*, CIE) defined alternative CMFs chosen such that any colour may be matched with positive amounts of three primaries which are imaginary stimuli existing only as mathematical constructs and are not physical realisable. These CMFs are denoted by $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ (CIE, 2004b) and shown in Figure 2-14.

Although there is individual variability in the properties of the human colour visual system, there is a need for a set of functions for colour specification which represents the mean properties of human observers with normal colour vision. Such a set of functions are termed the *Standard Colorimetric Observer*. The CIE has established two Standard Colorimetric Observers (details are given in the following sections) (CIE, 2004b).

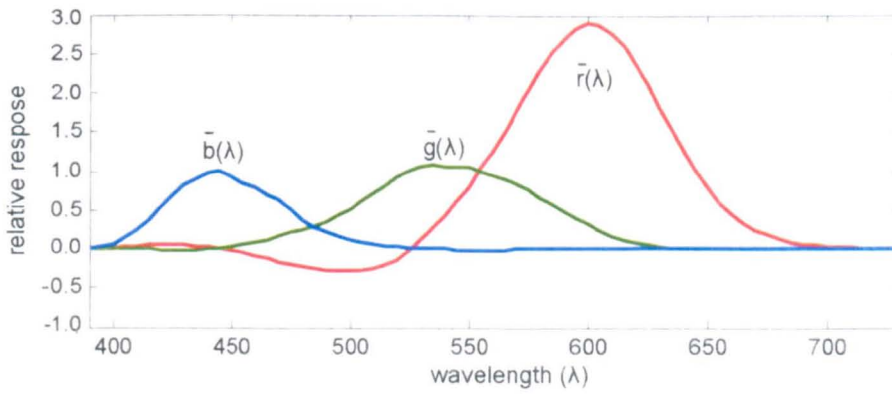


Figure 2-13: Colour matching functions: $\bar{r}(\lambda)$, $\bar{g}(\lambda)$ and $\bar{b}(\lambda)$ (CVRL, 2007).

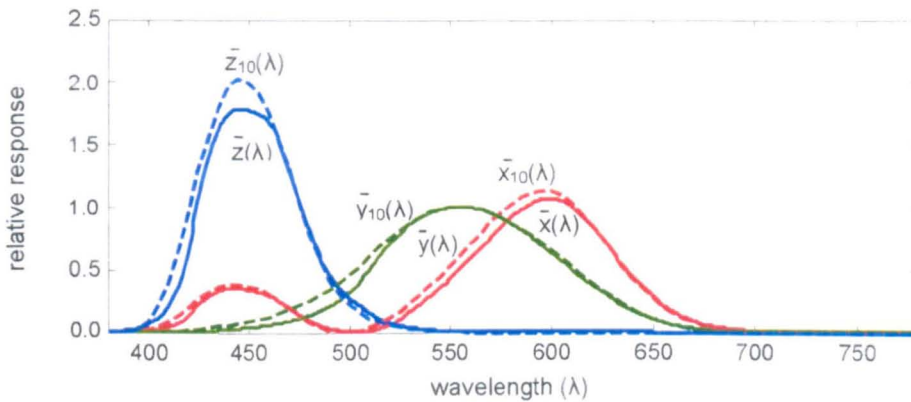


Figure 2-14: Colour matching functions: $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$: the CIE 1931 Standard Colorimetric Observer (solid line) and the CIE 1964 Standard Colorimetric Observer (dashed line) (CIE, 2004b).

2.3.1 CIE 1931 Standard Colorimetric Observer

The CIE 1931 Standard Colorimetric Observer (Figure 2-14) was adopted as the CMFs $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ of the standard observer for a 2° viewing field based on the colour matching experiments carried out by Guild (1931) and Wright (1928), and also the CIE 1924 luminous efficiency function $V(\lambda)$. Guild measured the CMFs using seven observers with broadband light as the primaries. Wright measured the CMFs of ten observers using monochromatic primaries at 650 nm, 530 nm and 460 nm. Both sets of data were transformed to a system that uses monochromatic primaries at wavelengths of 700.0 nm, 546.1 nm and 435.8 nm. It would have been possible to derive the CMFs if all the radiant powers of the monochromatic test stimuli had been recorded in the experiments. Instead, CMFs were derived by assuming that the luminous efficiency function $V(\lambda)$ is a linear combination of the CMFs $\bar{r}(\lambda)$, $\bar{g}(\lambda)$ and $\bar{b}(\lambda)$. Since the CMFs

$\bar{r}(\lambda)$, $\bar{g}(\lambda)$ and $\bar{b}(\lambda)$ had negative values at some wavelengths, which was inconvenient, they were transformed to the new set of CMFs $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ with the final adjustment of $\bar{y}(\lambda)$ to be the same as $V(\lambda)$.

2.3.2 CIE 1964 Standard Colorimetric Observer

The CIE recommended the 1964 Standard Colorimetric Observer (Figure 2-14) for a 10° viewing field based on data measured by Stiles and Burch (49 observers) and by Speranskaya (18 observers) (Stiles & Burch, 1959; Speranskaya, 1959). The CMFs $\bar{r}_{10}(\lambda)$, $\bar{g}_{10}(\lambda)$ and $\bar{b}_{10}(\lambda)$ were measured directly without using $V(\lambda)$. Stiles and Burch used different sets of primary stimuli (all monochromatic stimuli) for their investigation, but the final mean results were transformed to refer to primaries at 645.2 nm, 526.3 nm and 444.4 nm. Due to the presence of rods in 10° field, the luminance of the matching field was kept high in order to minimise the effect of rod participation in the results. Speranskaya used broadband primaries and the luminance levels of the experimental conditions were 30 to 40 times lower than in the Stiles and Burch study so that the results were affected by rod intrusion. In a manner similar to the CIE 1931 Standard Colorimetric Observer, these sets of CMFs $\bar{r}_{10}(\lambda)$, $\bar{g}_{10}(\lambda)$ and $\bar{b}_{10}(\lambda)$ were averaged, but weighted (the higher weight being given to Stiles and Burch study) after the correction for rod intrusion was made, and then, an all positive set of CMFs $\bar{x}_{10}(\lambda)$, $\bar{y}_{10}(\lambda)$ and $\bar{z}_{10}(\lambda)$ were derived. These CMFs are recommended for use when the viewing field is greater than 4°.

2.3.3 CIE XYZ Tristimulus Values

The CIE XYZ tristimulus values X , Y and Z are defined from the CIE Standard Colorimetric Observer CMFs by following Equations 2-4 (CIE, 2004b).

$$\begin{aligned}
 X &= k \sum_{\lambda=\lambda_{\min}}^{\lambda_{\max}} Q(\lambda) \bar{x}(\lambda) \Delta\lambda \\
 Y &= k \sum_{\lambda=\lambda_{\min}}^{\lambda_{\max}} Q(\lambda) \bar{y}(\lambda) \Delta\lambda \\
 Z &= k \sum_{\lambda=\lambda_{\min}}^{\lambda_{\max}} Q(\lambda) \bar{z}(\lambda) \Delta\lambda
 \end{aligned}
 \tag{Equation 2-4}$$

where $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ are the CMFs of the CIE 1931 Standard Colorimetric Observer, λ_{min} and λ_{max} are the minimum and maximum wavelengths of the range over which the spectral power distribution is sampled (usually 360-830 nm or 380-780 nm) and $\Delta\lambda$ represents the wavelength sampling interval. For a reflecting or transmitting object, $Q(\lambda)$ can be replaced by $S(\lambda)O(\lambda)$, where $S(\lambda)$ is the spectral power distribution of the light illuminating the object and $O(\lambda)$ is the spectral reflectance factor or the spectral transmittance factor of the object. k in the above equation is a factor used to normalize the tristimulus values such that a perfectly reflecting or transmitting object will have $Y = 100$; it can be expressed as Equation 2-5.

$$k = \frac{100}{\sum_{\lambda=\lambda_{min}}^{\lambda_{max}} S(\lambda)\bar{y}(\lambda)\Delta\lambda} \quad \text{Equation 2-5}$$

For a self-luminous object or illumination, $Q(\lambda)$ is the spectral power distribution and k is an arbitrary normalising factor, which is usually set to 683 lm/W so that Y corresponds to the luminance in cd/m^2 .

The CIE 1964 XYZ tristimulus values can be calculated in a similar manner using the CIE 1964 CMFs of the Standard Colorimetric Observer $\bar{x}_{10}(\lambda)$, $\bar{y}_{10}(\lambda)$ and $\bar{z}_{10}(\lambda)$ instead of the $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$. However, unlike $\bar{y}(\lambda)$, $\bar{y}_{10}(\lambda)$ is not adjusted to the luminous efficiency so Y value does not represent the luminance in cd/m^2 .

2.3.4 CIE xy Chromaticity Diagram

As described above, the colour of a stimulus can be specified using the CIE XYZ tristimulus values. *Chromaticity coordinates* (x , y and z) (CIE, 2004b) are defined as the ratio of the tristimulus values to their sum as shown in Equation 2-6.

$$x = \frac{X}{X+Y+Z}, \quad y = \frac{Y}{X+Y+Z}, \quad z = \frac{Z}{X+Y+Z} = 1 - x - y \quad \text{Equation 2-6}$$

Chromaticity coordinates can be plotted to give a chromaticity diagram which is a two dimensional representation of colours. All monochromatic wavelengths map to a position along the boundary, which is called the *spectral locus*. The CIE xy chromaticity diagrams of the CIE 1931 and the CIE 1964 Standard Colorimetric Observers are shown in Figure 2-15 (a). Chromaticity coordinates are relative so that

colours which have the same relative spectral power distribution map to the same chromaticity coordinates. Since the chromaticity diagram follows the properties of additive colour mixture, as shown in Figure 2-15 (b), the chromaticity point of an additive mixture of αA and βB ($\alpha \geq 0$, $\beta \geq 0$ and $\alpha + \beta = 1$) falls on the straight-line connecting the chromaticity points of A and B .

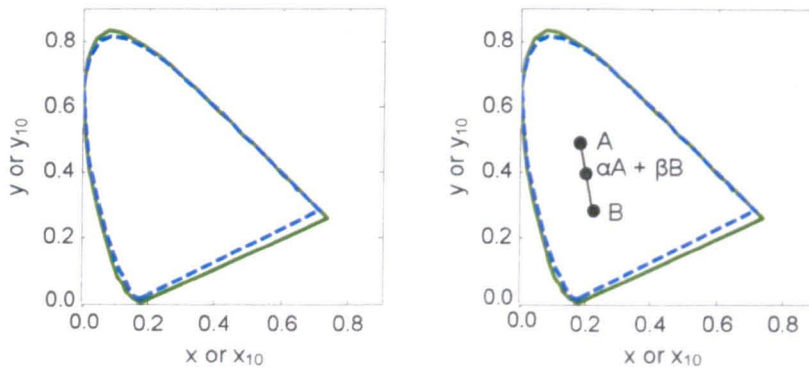


Figure 2-15: *xy* chromaticity diagrams of the CIE Standard Colorimetric Observer. (a) The CIE 1931 Standard Colorimetric Observer (solid line) and the CIE 1964 Standard Colorimetric Observer (dashed line). (b) Chromaticity diagram with the position of the additive mixture of $\alpha A + \beta B$.

2.3.5 CIE 1976 UCS Diagram

Although, the CIE *xy* chromaticity diagram is very useful for quantifying colour stimuli, it has one serious disadvantage which is the non-uniformity of colour distribution within its space. Equal changes in chromaticity coordinates do not correspond to equal perceptual differences. A different chromaticity diagram known as the *CIE 1976 Uniform Chromaticity Scale (UCS) diagram* was developed by attempting to make a chromaticity diagrams that gave a more perceptually uniform representation of colour differences (CIE, 2004b). The chromaticity diagram is produced by plotting u' and v' which are defined by Equation 2-7.

$$u' = \frac{4X}{X + 15Y + 3Z} = \frac{4x}{-2x + 12y + 3}$$

$$v' = \frac{9Y}{X + 15Y + 3Z} = \frac{9y}{-2x + 12y + 3}$$

Equation 2-7

2.3.6 CIELAB Colour Space and Colour Differences

A further attempt to define a perceptually uniform colour space is the *CIELAB colour space* defined by the CIE in 1976 (CIE, 2004b). A colour stimulus can be described using perceptual attributes such as lightness, chroma and hue. The CIELAB system is comprised of three orthogonal dimensions L^* , a^* and b^* . The vertical dimension L^* represents the lightness; the two horizontal dimensions a^* and b^* represent the redness-greenness and yellowness-blueness perceptions of colours respectively. These coordinates are defined by the following transformation of the CIE XYZ tristimulus values given in Equation 2-8.

$$\begin{aligned}
 L^* &= 116f(Y/Y_n) - 16 \\
 a^* &= 500[f(X/X_n) - f(Y/Y_n)] \\
 b^* &= 200[f(Y/Y_n) - f(Z/Z_n)]
 \end{aligned}
 \tag{Equation 2-8}$$

$$f(\omega) = \begin{cases} (\omega)^{1/3} & \text{if } \omega > (24/116)^3 \\ (841/108)(\omega) + 16/116 & \text{if } \omega \leq (24/116)^3 \end{cases}$$

where X , Y , Z and X_n , Y_n , Z_n are the CIE XYZ tristimulus values for the stimulus and for the reference white respectively. The chroma C_{ab}^* and hue angle h_{ab} can be calculated from the following Equation 2-9.

$$\begin{aligned}
 C_{ab}^* &= \sqrt{a^{*2} + b^{*2}} \\
 h_{ab} &= \tan^{-1}(b^*/a^*)
 \end{aligned}
 \tag{Equation 2-9}$$

Practical application of colorimetry requires the ability to evaluate a colour difference between two colour stimuli, which corresponds to the magnitude of the perceptual difference of two colour stimuli. The colour difference specified by CIELAB is quantified as the Euclidean distance between the co-ordinates of the two stimuli. These distances are expressed in Equations 2-10.

$$\begin{aligned}
 \Delta E_{ab}^* &= \sqrt{\Delta L^{*2} + \Delta a^{*2} + \Delta b^{*2}} = \sqrt{\Delta L^{*2} + \Delta C^{*2} + \Delta H^{*2}} \\
 \text{where } \Delta H_{ab}^* &= 2(C_{ab,1}^* C_{ab,2}^*)^{1/2} \sin[(h_{ab,2} - h_{ab,1})/2]
 \end{aligned}
 \tag{Equation 2-10}$$

where ΔE_{ab}^* is the *CIELAB colour difference* and the subscripted number 1 and 2 indicates the two colours whose colour difference is to be compared.

2.3.7 S-CIELAB Colour Difference

S-CIELAB (Zhang & Wandell, 1997) has been proposed as a spatial extension to CIELAB to account for the influence of spatial information on the colour appearance and colour discrimination of colours in digital images. The development of the CIELAB system was based on psychophysical data from the colour appearance assessments of relatively large uniform fields (subtend at least a 2° viewing field). Therefore, CIELAB was found to be unsatisfactory when predicting the colour appearance and discrimination of smaller fields or fine patterned colours. For example, when a continuous-tone colored image is compared with a halftone version of the image, a point-by-point computation of the CIELAB colour difference produces large errors at most image points. Because the halftone patterns vary rapidly in a spatial sense, perceived these differences are blurred and the reproduction may still preserve the appearance of the original.

A flowchart representing the computational procedure of S-CIELAB is given in Figure 2-16. The values representing the input image are initially converted into opponent-colour values that represent a luminance (achromatic), red-green yellow-blue channels. Spatial filtering is then performed to the channels individually. The filters are selected with regard to the sensitivity of each channel of the human visual system. The final output can be computed in a manner similar to CIELAB ΔE^*_{ab} from the CIE XYZ values transformed from the filtered images.

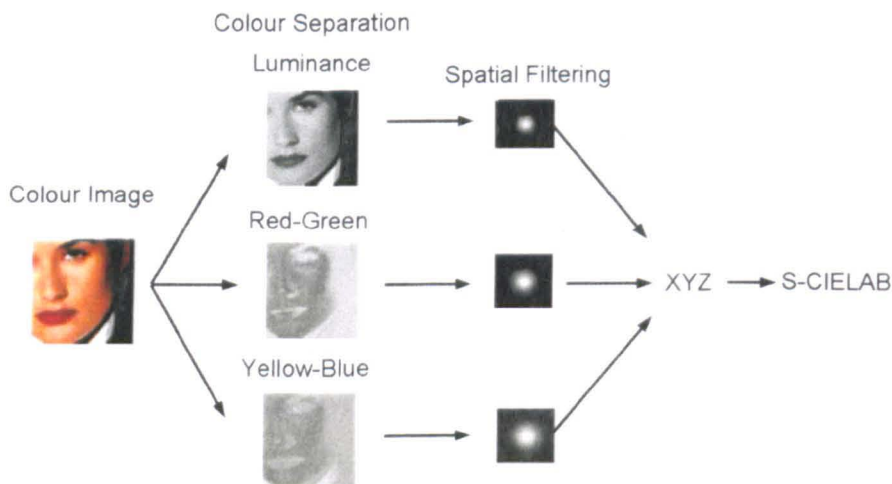


Figure 2-16: A flowchart of S-CIELAB computational process (Zhang & Wandell, 1997).

2.4 Light-Field Formation and Measurements

The human visual system is activated by light entering into the eye. Light radiated from light sources is reflected, refracted, scattered or diffused by the surface of an object. Interaction of all these sources of radiation creates a spatially and temporally distributed light-field that we perceive. It is important to understand the behaviour of light in terms of this interaction with a surface and the methods of instrumental measurement of light or colour.

2.4.1 Light Behaviour

When light is incident on the surface of an object, two phenomena occur. One is reflection and the other is transmission as shown in Figure 2-17.

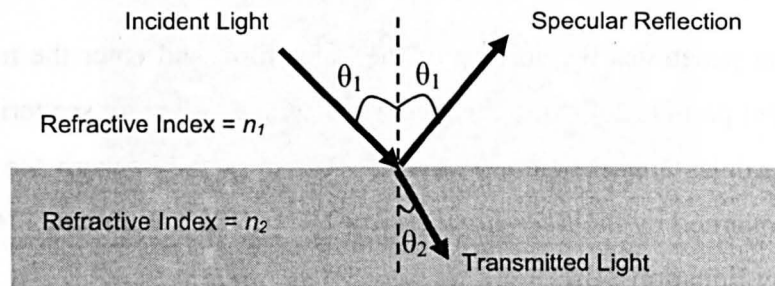


Figure 2-17: Relationship between incident light, specular reflection and transmitted light.

The relationship between the incident and refracted light can be stated according to *Snell's Law*. When light passes through a medium of refractive index n_1 and enters a medium of refractive index n_2 , the light is bent (refracted) through an angle (Figure 2-17) which can be expressed by Equation 2-11.

$$\frac{\sin \theta_1}{\sin \theta_2} = \frac{n_2}{n_1} \quad \text{Equation 2-11}$$

where θ_1 and θ_2 are the angles of the incident and refracted light respectively. The refractive index is dependent on the wavelength of light used.

The light reflected from the surface of, for example, a paint film without transmission into the body of the medium usually has the same colour as the incident light, when it is viewed at an angle equal to that of the incident light θ_1 ; this directly reflected light is called the *specular reflection* (Figure 2-17). The angle of the specular

reflection follows the *law of reflection* which states that the angle between the incident light and the normal (the perpendicular to a surface) is equal to the angle between the specular reflection and the normal. The amount of specular reflectance R and transmittance T for unpolarised light can be described by the angles of the incident light θ_1 and refracted light θ_2 determined from *Fresnel's law*, Equation 2-12.

$$R = \frac{1}{2} \left[\frac{\sin^2(\theta_1 - \theta_2)}{\sin^2(\theta_1 + \theta_2)} + \frac{\tan^2(\theta_1 - \theta_2)}{\tan^2(\theta_1 + \theta_2)} \right] \quad \text{Equation 2-12}$$

$$T = 1 - R$$

When the light enters a medium with an angle of incidence normal to the surface (0°), the specular reflection R can be defined by the relative refractive indices of the two media n_1 , and n_2 , as expressed in Equation 2-13.

$$R = \left(\frac{n_2 - n_1}{n_2 + n_1} \right)^2 \quad \text{Equation 2-13}$$

Light then penetrates the surface of the paint film, and enter the medium which contains pigment particles of many different sizes with a variety of scattering properties. When these particles are sufficiently small ($< \sim 0.2 \mu\text{m}$), they absorb the light without scattering as governed by the *Beer-Lambert law* (Beer, 1852; Lambert, 1760) which can be expressed by Equation 2-14.

$$I = I_0 10^{-\epsilon c l} \quad \text{Equation 2-14}$$

where I and I_0 are the intensity of the transmitted and incident light respectively, ϵ is the molar absorption coefficient ($1 \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$), c is the concentration of the absorbing species in the medium ($\text{mol} \cdot \text{l}^{-1}$) and l is the distance that the light travels through the layer (the path length, cm). When it is described as $T = I / I_0$, Equation 2-14 can be written as Equation 2-15.

$$\log(I / T) = \epsilon c l = A \quad \text{Equation 2-15}$$

where T represents the transmittance and A is the absorbance. Since absorbance follows the additivity law, if the particles have different absorbance values, the sum of these gives the resultant absorbance. However, the Beer-Lambert law is restricted to a mono-molecular pigment condition and application to non-scattering pigment particles. The *Kubelka-Munk theory* (Kubelka & Munk, 1931) would be more appropriate, which can be applied to translucent objects containing pigment particle distributions that are reasonably uniform.

When the pigment particles are larger, with a diameter in excess of ten times the wavelength, the light is not only absorbed but also scattered. The amount of scattering depends on the refractive indices of the pigments and the medium following *Fresnel's law*. The scattered light therefore depends on the changes of the wavelength of the light. As shown in Figure 2-18, the light in the layers is scattered between the pigment particles or absorbed by these particles in many different directions due to the irregular size and shape of the particles. Parts of the scattered light reach the top surface and this light then leaves the surface at all angles. This produces *diffuse reflection*.

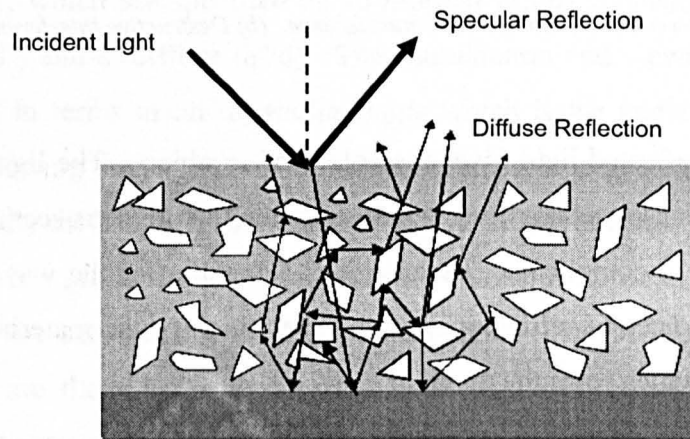


Figure 2-18: Microscopic view of light travelling through a layer.

Another important phenomenon is *interference*. This involves the interaction between waves of light as illustrated in Figure 2-19. When two light waves arrive at the same place, they add together. If the two waves are in phase (the peak of one wave coincides with the peak of another wave), the resulting amplitude is doubled. This is called *constructive interference* (Figure 2-19 (a)). If two waves are out of phase (the peak of one wave coincides with the trough of another), they cancel each other out. This is called *destructive interference* (Figure 2-19 (b)). Since the light consists of waves of various wavelengths, the reinforced visible wavelengths are pronounced more strongly in the perceived colour.

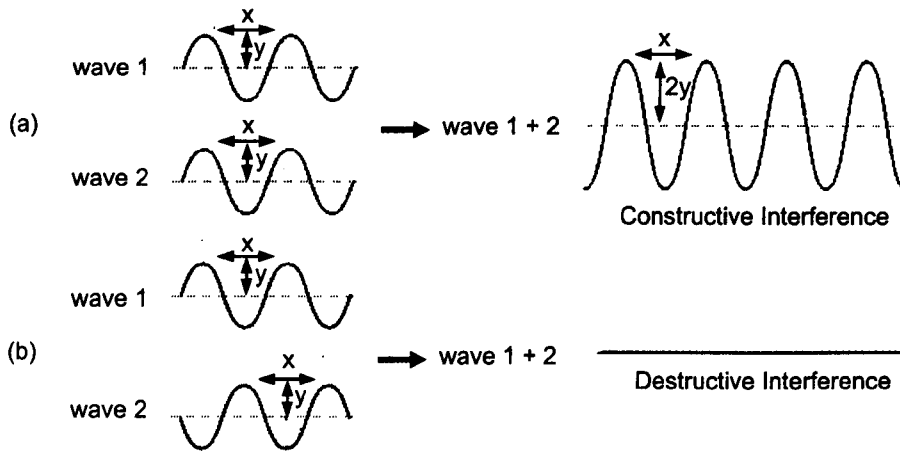


Figure 2-19: Light interference. (a) Constructive interference. (b) Destructive interference.

The resulting reflected light gives the colour of an object. The light behaviour is basically a simple concept, as has been demonstrated in the previous sections. However, the total reflectance is complicated, because it is a function of the wavelength of the incident light, the indices of refraction, and the absorption of the material and also the size of the particles which form the material.

Characteristics of the reflected light in accordance with the surface roughness or smoothness are generally illustrated in Figure 2-20. When the light illuminates a mirror-like surface (a smooth polished surface), the incident light entirely reflects at the specular angle (Figure 2-20 (a)). If the surface is not completely polished, but not completely matt, the light is reflected in many directions but with a lower specular reflection (Figure 2-20 (b)). The light indiscriminately reflects at all angles without any specular reflection (diffuse reflection) when the surface is completely matt (Figure 2-20 (c)).

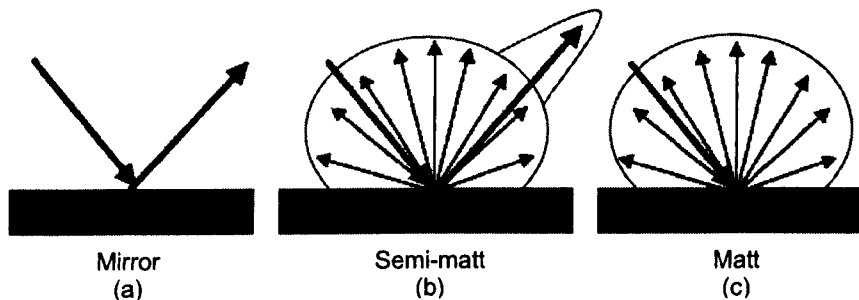


Figure 2-20: Light reflection of three types of surface. (a) Mirror. (b) Semi-matt. (c) Matt.

2.4.2 Light/Colour Measurements

Colour measuring instruments, such as a spectroradiometer and a spectrophotometer, are designed to measure colours in terms of radiance, reflectance and the CIE colorimetric specifications. As has been demonstrated in the previous section, the angles of illumination and viewing are important. Therefore, in colour measuring instruments, the illumination and viewing geometry are critical. The CIE has recommended different types of geometry which are appropriate for the measurement of surface colours (CIE, 2004b). Schematic diagrams of four of these geometries are given in Figure 2-21, which are specified as 45°/normal (45/0) and Normal/45° (0/45), and Diffuse/8° (d/8°) and 8°/diffuse (8°/d). The illumination and viewing geometry can be also described in terms of an *aspecular angle* which is the angle between a specular angle and a viewing angle. For colour measurement, it is generally recommended to only measure the diffuse reflection and not the specular reflection. Only the diffusely reflected light has properties that represent the colour of an object, since the specular reflection has the same colour as the incident light. The 45/0 and 0/45 geometries (Figure 2-21) are therefore recommended so as to avoid the specular reflection. A diffusely illuminating white sphere, called an integrating sphere, is used to produce the d/8 and 8/d geometries. As illustrated in Figure 2-21, this is able to control whether the specular component is included or excluded in the measurement by adjusting a gloss trap mounted at the specular angle on the integrating sphere. If the reflection from the surface is perfectly diffused, the difference between the illumination and viewing geometries does not affect the colour. However, since most surfaces of materials are not perfectly matt, the geometry has to be specified. Some materials, such as metallic and pearlescent coatings, are especially dependent on illumination and viewing geometry. Therefore, the measurements made using more than two geometries are necessary to characterise their surface properties.

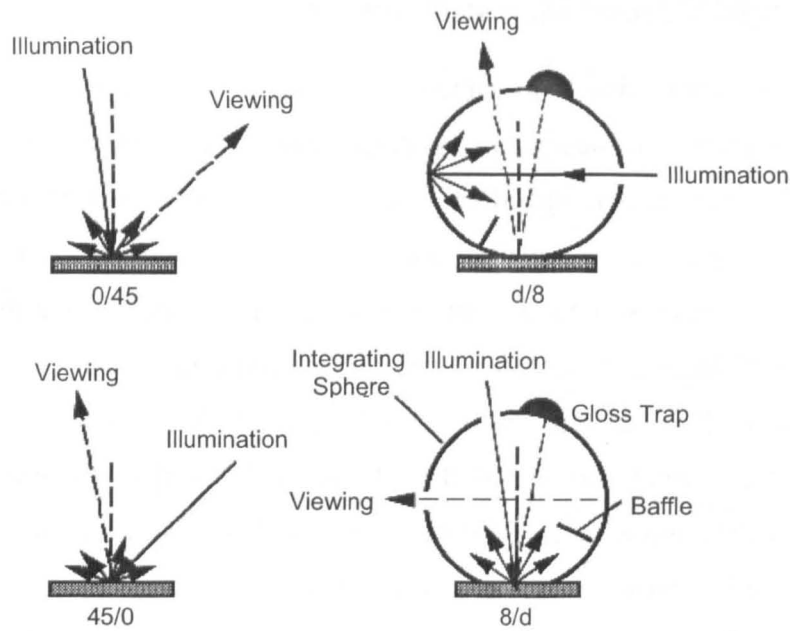


Figure 2-21: Schematic diagrams showing the CIE recommended illumination and viewing geometries: left 0/45 and 45/0, right d/8 and 8/d (Luo, 2002).

2.4.2.1 Spectroradiometer

A spectroradiometer is an instrument designed to measure the radiometric quantities, for example irradiance (in $\text{W}\cdot\text{m}^{-2}$ unit) or radiance (in $\text{W}\cdot\text{sr}^{-1}\text{m}^{-2}$ unit), of an object over the visible spectrum with a fixed measurement interval such as 5nm or 10nm. A tele-spectroradiometer (TSR) is the most frequently used type of spectroradiometer. This normally consists of a telescope, a monochromator and a photo-detector and can be used for measurements of both surface and self-luminous colours. For accurate measurement, the instrument needs to be calibrated which is usually done by using a standard light source whose absolute spectral power distribution is known. Such standard light sources are usually available from national standardising laboratories (*e.g.*, National Physical Laboratory) (Hunt, 1998). The advantage of a TSR is that the resulting measurements can correspond to the actual conditions of viewing (Hunt, 1998). This can be achieved by setting up a TSR at the same position as was occupied by an observer's eye, and by directing it at a target colour while it is illuminated by the same illumination in the same surroundings.

2.4.2.2 Spectrophotometer

A spectrophotometer is often used to make measurements of surface colours. A spectrophotometer is designed to measure the ratio between the incident and reflected light of a surface across the visible spectrum with a fixed interval such as 5 nm, 10 nm or 20 nm, to give a reflectance value of a surface colour. The basic elements of the instrument are a light source, a monochromator and a detector. The instrument can be calibrated by comparison with the data of a set of standard samples, the BCRA-NPL ceramic tiles which are available from the National Physical Laboratory for this purpose. One of the CIE recommended geometries (Figure 2-21) is usually employed as illumination and viewing geometry of the instruments. Some spectrophotometers, often called gonio-spectrophotometers, are able to make measurements at a wide range of illumination and viewing geometries. This is particularly useful for the measurements of surfaces whose colours change as a function of these angles.

2.5 Metallic Coatings

Ever since the first automobiles were made in the late 1800s, there have been many changes in paint technologies. The majority of automotive paints produced until the 1950s were *solid-colour coatings* which are coatings with a uniform appearance irrespective of the angle of illumination and viewing. Later, *special-effect coatings* such as *metallic* and *pearlescent coatings* were introduced (McCamy, 1996). These special-effect coatings exhibit differences in their perceived appearance with changes in illumination and viewing angle and thus these coatings are sometimes called *gonioapparent* coatings (Rodrigues, 2004) The increase in popularity of these coatings is due to the glamorous look which they give to a car. These coatings are also widely used on other products such as bikes, home electrical products and sportswear.

The metallic coatings contain aluminium pigments which possess many of the characteristics of aluminium metal itself. Among these are resistance to corrosion, light reflection (visible, infrared and ultraviolet) and a unique lamellar geometry to the individual particle. The aluminium pigments are mainly manufactured from over 90 % pure aluminium metal, mineral spirits and suitable fatty acid, generally stearic or oleic. These three materials are atomised using a ball mill and flattened into flakes with a typical thickness between 0.1 and 2 μm and a diameter of between 5 and 50 μm . Thus,

the aluminium pigments are actually flakes, rather than being spherical or granular as are most other pigments (Akzo Nobel, 2004).

An automobile coating is usually comprised of three layers as shown in Figure 2-22 (Akzo Nobel, 2004). The *ground coat* is applied on to the substrate (panel). Then the *colour coat (base coat)* is treated. There are different types of the colour coat. A plain colour coat only contains solid-colour paint. A colour coat with special effects contains a mixture of solid-colour paint and metallic paint, or a mixture of solid-colour paint and pearlescent paint. On top of these two coatings, the *clear coat* is applied in order to protect the lower layers against weathering: it also gives the high gloss properties associated with these coatings.

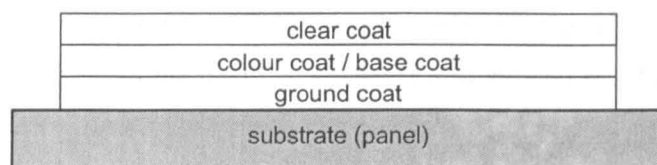


Figure 2-22: Coating layers on an automotive panel.

When metallic paint is incorporated within a coating, the aluminium flakes exist in many layers and do not tend to lie parallel to the coating surface (but this is dependent on the systems used). As the coating dries the layer becomes thinner and the flakes become more parallel to the surface. The orientation of aluminium flakes is one factor that gives the unique appearance of metallic coatings, since it affects the behaviour of reflected light (see Section 2.4.1). As shown in Figure 2-23 (a), the parallel alignment of the flakes allows more light to be reflected in the specular direction (mirror-like reflection) and thus gives less diffuse reflection. In addition, it leads to the appearance changes with the viewing angle. In contrast, as shown in Figure 2-23 (b), the light reflected from the randomly oriented flakes is scattered diffusely (Besold, 1990).

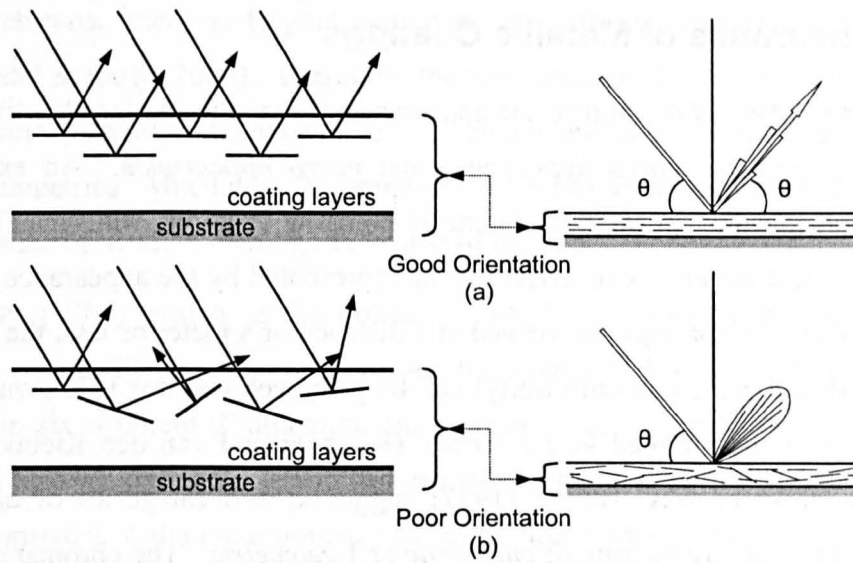


Figure 2-23: Light behaviour caused by the orientations of aluminium flakes. (a) Parallel alignment of the aluminium flakes. (b) Random alignment of the aluminium flakes.

The appearance of the coatings is not only influenced by the orientation of the flakes, but also their size. When the flakes are oriented parallel to the surface, the light reflects at the surface of the flakes and comes back in a specular direction, but the light reflected at the edge is scattered and comes back to the surface as diffusely reflected light. An example of the light reflected from coatings with coarse and fine flakes is given in Figure 2-24; a panel coated with coarse flakes leads to high brilliance or sparkle because of the more mirror-like reflection properties due to the larger area of the flakes, whereas a panel coated with fine flakes is much smoother but also darker due to the more diffuse reflection (Akzo Nobel, 2004). The arrows denote mirror-like reflection and the asterisks indicate diffuse reflection in Figure 2-24.

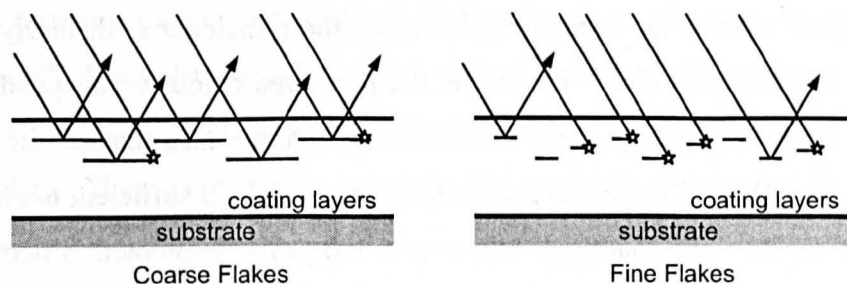


Figure 2-24: Relationship between aluminium flake size and light reflection.

2.5.1 Appearance of Metallic Coatings

McCamy (1996, 1998) studied the appearance of metallic coatings by dividing the subject into two categories: *macro appearance* and *micro appearance*. An example of macro appearance is the appearance of a metallic-coating material, which can be seen at a distance of a few meters. This effect can be represented by the appearance of colour and gloss. When the coatings are viewed at a distance of a meter or less, the details of the coatings (small-scale non-uniformity) can be perceived and this is known as micro appearance. It is also termed *visual texture* (Kirchner and van den Kieboom *et al.*, 2007). In a different way, Hunter (1977) suggested two categories of appearance criteria that are normally measured: *chromatic* and *geometric*. The chromatic attribute is the colour of a surface whereas the attributes associated with the geometric category could include gloss, haze, texture, *etc.* Thus, McCamy and Hunter grouped appearance attributes differently. However, they both stress the importance of texture which is the focus of the present study. In the thesis, the term *visual texture* is used henceforth.

The following sections focus on the essential attribute, colour and the target attributes of this study, the texture of metallic coatings.

2.5.1.1 Colour of Metallic Coatings

Colour is obviously one of the most important attributes when describing the appearance of coatings (van Aken, 2006). The CIE introduced methods to quantify colours from the spectral properties of light reflected from a coating surface measured using a spectrophotometer (or a gonio-spectrophotometer) or spectroradiometer (or a gonio-spectroradiometer) (see Section 2.3 and Section 2.4). Traditional solid-colour pigments absorb part of the incident light, while the remainder is diffusely scattered (Teaney and Pfaff *et al.*, 1999). Therefore, the perceived colour is independent of the illumination and viewing angle and, consequently, a single instrumental measurement using one of the CIE recommended geometries such as 0/45 is sufficient to characterise the colour of solid-colour coatings. In contrast, the primary interaction between light and metallic coatings is the specular reflection from the aluminium flakes. Thus, the perceived brightness depends on the viewing angle, but is independent of the illumination angle, while the perceived chroma and hue are independent of both the illumination and viewing geometry (Nadal & Early, 2004). The pearlescent coatings usually contain mica interference pigments (interference; see Section 2.4.1) so that their

perceived chroma, hue, and brightness depend on both the illumination and viewing angles (Nadal & Early, 2004). Therefore, the wide angular dispersion of reflected light in these gonioapparent materials can be fully characterised by evaluating them using a range of geometries. Much work has been carried out to try to measure and characterise the appearance of metallic coatings and some of this work is reviewed in the followings.

Alman (1984) evaluated the colour of metallic coatings instrumentally. The reflectance of metallic coatings with various flake types and solid-colour pigments was measured in six different illumination and viewing (detector position) geometries. It was found that an optimum set of geometries to characterise the colour of metallic coatings consisted of the measurements taken at three viewing angles. The angles are near the specular angle (aspecular angle of 15°), about 45° from the specular angle (aspecular angle of 45°) and far from the specular angle (aspecular angle of 110°) with an illumination angle of 45° from normal to the surface of a metallic coating. It was determined that the colour could be characterised by an interpolation between the minimum of these three measurements in terms of CIELAB L^* using a second order polynomial; more measurements would only be served to improve the precision of the same function. In addition, a comparison was made between measurements made using varying illumination angles and a fixed viewing angle, and measurements made using a number of viewing angles with a fixed angle of illumination. It was found the angular dependence of the CIELAB values for a typical metallic colour was not influenced by these geometric differences. Later, the importance of these three angles (aspecular angles of 15° , 45° and 110°) were verified by Rodrigues (Westlund & Meyer, 2001).

Venable (1987) proposed a simple theoretical model for the reflectance factors as a function of a viewing angle, and was also found that measurements at three angles adequately characterise metallic coatings. The angles should be near the specular angle, more than 60° from the specular angle and an intermediate angle between them. However, slightly different geometries from the studies previously introduced, were recommended. They were measurements at the aspecular angles of 20° , 40° and 75° with normal incident illumination.

Saris and Gottenbos *et al.*(1990) studied correlations between colour differences perceived by human observers and colour differences based on instrumental measurements. Instrumental measurements at seventeen different geometries (from three different instruments) were compared with the associated perceptual colour differences. It was found that the perceptual colour differences could be best

characterised using the instrumental measurements at the aspecular angles of 25°, 45° and 110° with an illumination angle of 45° from normal.

Nicholls (2000), in a study similar to that of Saris and Gottenbos *et al.*, also investigated correlations between instrumental and visual measurements. Instrumental measurements were made using a spectrophotometer with aspecular angles of 15°, 25°, 45°, 75° and 110° with an illumination angle of 45° from normal. Additionally, using an integrating sphere spectrophotometer, d/8 geometry with the specular component included (SPIN) and excluded (SPEX) were used. A total of 35 observers, consisting of 17 professional (experienced) observers and 18 naïve (student/untrained) observers, carried out the visual assessment of the colour difference between pairs of metallic coatings. A statistical analysis of the results showed that there were significant differences in performance between the professional and naïve observer groups; the professional observers were more consistent in their judgements. The measurement angles that were found to be important to describe the perceptual colour differences from the professional observers were SPIN, 25°, 45° and 110° aspecular. In comparison with this distinct group, the combined group of the professional and naïve observers was considered as a representation of the general population. The angles of SPIN, 15°, 45° and 110° aspecular were shown to be important angles for the general population.

Chou (2003) also evaluated the colour difference of metallic coatings in terms of instrumental and visual measurements. Instrumental measurements were made at aspecular angles of 20°, 45°, 75° and 110° with an illumination angle of 45° from normal. The results showed that the combination of the measurements at the aspecular angles of 15°, 45° and 110° could characterise the perceptual colour difference. This indicates that the measurement at the aspecular angle of 75° could be considered redundant.

Various sets of the geometry have been recommended for the measurement of the colour of metallic coatings by different researchers. A common finding from these studies is the necessity for measurement at the more than one geometry for metallic coatings, and most researchers recommend at least three angles; one near the specular angle; a second at around 45° from the specular angle; and a third at far from the specular angle, with an illumination angle of 45° from the normal.

The American Society for Testing and Materials (ASTM) (2003) recommended aspecular angles of 15°, 45° and 110° with an illumination angle of 45° from the normal, while Deutsches Institut für Normung (DIN) has recommended aspecular angles of 25°, 45° and 75° with an illumination angle of 45° from normal (Imura, 2006).

A more intensive approach to the characterisation of macro appearance is to measure the bidirectional reflectance distribution function (BRDF) which describes the geometrical reflectance properties of the surface in absolute terms. Computational models have been proposed to predict the BRDF from multiple scattering from microstructural properties (*e.g.*, surface roughness, pigment size, and pigment spatial distribution) (Sung and Nadal *et al.*, 2000; Sung and Nadal *et al.*, 2002). The BRDF gives more information about the distribution of the reflected light which can be utilised to correctly render images for appearance simulation.

2.5.1.2 Visual Texture of Metallic Coatings

There is a large amount of literature on texture in general which will be introduced in the next section. The importance of visual texture for metallic coatings, has been referred to, but the available literature is relatively limited, *e.g.*, (Kang and Butler *et al.*, 2000; Kirchner and van den Kieboom *et al.*, 2007; McCamy, 1998). It is known that visual texture affects the perception for colours (Han and Luo *et al.*, 2005a, b; Hunter, 1987). Unlike colour, there is no standardised method for numerical evaluation of visual texture. Therefore, in order to model total appearance, visual texture needs to be perceptually and instrumentally (computationally) characterised.

Visual texture changes depending on illumination and viewing geometry, surface roughness, and the size and spatial orientation of aluminium flakes, as well as other appearance attributes of metallic coatings.

In the scientific study of visual texture for metallic and pearlescent coatings, one particular aspect of appearance is often considered. This has been called glint (Kirchner and van den Kieboom *et al.*, 2007; McCamy, 1998), sparkle (Baba and Miura *et al.*, 2005; Đurikovič, 2003; Ershov and Kolchin *et al.*, 2001; Prakash and Karmes *et al.*, 2005; Rodrigues, 2004) or glitter (Arai & Baba, 2005; ASTM, 2006; McCamy, 1998; Sung and Nadal *et al.*, 2000). According to the Cambridge English Dictionary (Cambridge, 2005), general definitions of these words are “small bright flashes of light reflected from a surface” for glint, “bright shine with a lot of small points of light” for sparkle and “a lot of small bright flashes of reflected light” for glitter. So the general definitions of these words are, in fact, very similar. However they may be used to indicate different phenomenon. For instance, glitter is also applied bright areas which are seen under directional illumination but it should not be confused with tiny spots or

sparkle caused by individual flakes (Đurikovič, 2003). Even if these three terms imply the same phenomenon (such as bright flashes/spots), they are made complicated by the gonioapparent nature of metallic coatings (McCamy, 1998). For example, high intensity bright flashes/points can be seen on metallic coatings at small aspecular angles where a large percentage of the flakes reflect light specularly, but the contrast between the bright flashes/points and their background is not as high as it is at large aspecular angles. At large aspecular angles bright flashes/points are more separate and distinct from the background due to the high contrast between them. This indicates that at least two different phenomena cause the attributes termed glint, sparkle and glitter.

Another important aspect of visual texture for metallic coatings is called coarseness (Kirchner and van den Kieboom *et al.*, 2007) or graininess (McCamy, 1998). For example, coarseness can be described as “composed of relatively large parts or particles (Merriam-Webster, 2005)”, “loose or rough in texture (Merriam-Webster, 2005)” or “related to the spatial repetition period of the local structure (Xin and Shen *et al.*, 2005)”. These terms seem to relate to the most fundamental effect of visual texture for many materials (Shen & Bie, 1992; Tamura and Mori *et al.*, 1978; Xin and Shen *et al.*, 2005). Although these attributes are very common for describing texture, again their appearance for metallic coatings changes due to goniometric effects. McCamy (1998) remarked that when a metallic coating is viewed at small and large aspecular angles, the bright flashes/points can be observed, but at intermediate angles, the attribute turns to graininess.

It has been seen that, although general concepts of common visual texture attributes are implicitly simple, they do not clearly indicate the phenomenon which can be seen on metallic coatings because of the complexity of the appearance of the coatings. This suggests that it is necessary to define the attributes in such a manner as to include a specification of the viewing conditions.

The sponsor of this project, Akzo Nobel identifies the terms glint and coarseness as the most important visual texture attributes of metallic coatings and, in order to avoid confusion, proposed strict definitions taking into account the viewing conditions (Kirchner and van den Kieboom *et al.*, 2007):

“Glint Impression is the overall impression of several or many tiny light-spots (“glints”) that are strikingly brighter than their surrounding. Instead of brightness, it may also be their color that distinguishes the glints from the background. The glints may be expected to switch on and off when the observation/illumination geometry is changed. Glint Impression is visible under intense unidirectional illumination conditions only.”

“Diffuse Coarseness is the perceived contrast in the light/dark irregular pattern exhibited by effect coatings viewed under diffuse illumination conditions.”

There are other attributes of visual texture which can be seen on metallic coatings such as coherence glitter and mottle (McCamy, 1998). However, they are more difficult to recognise visually and therefore considered insignificant to the current state of appearance modelling.

Instrumental measurements of macro appearance, e.g., colour, are often made by macroscopic methods using a spectrophotometer or a spectroradiometer. However, generally these instruments give the information from integration of an area which is too large to resolve the necessary detail for texture analysis. Therefore, microscopic methods, using a CCD camera or a micro-spectrophotometer must be employed to analyse texture information. For texture in general, a variety of computational methods have been proposed in order to describe texture numerically, based on information obtained from two-dimensional digital images (this will be reviewed in the following sections). Although many texture analysis methods exist, there is no recognised method for any surface. Therefore, it is necessary to develop a method for numerical expression of texture, particularly for metallic coatings.

A study by Arai & Baba (2005) investigated the texture of metallic and pearlescent coatings based on a two-dimensional digital image captured using a camera with a geometry of 45/0. Sets of silver, blue and red metallic coatings were used as samples where each colour group had four samples that were different in the physical size of the aluminium flakes contained in the coatings: fine, medium fine, medium and coarse. It was found that a power spectrum derived from the Fourier transform of the

two-dimensional brightness distribution is larger when the coatings contain larger sizes of flakes (see Section 2.6.2.1). Although this implies a relation between visual texture and the spatial frequency domain, the extracted information from an image was evaluated according to the physical compositions of the coatings and not compared with the perceptual coarseness or fineness.

It should be noted that manufacturers use the term coarse or fine to indicate the physical size of aluminium flakes contained in coatings. Perceptual coarseness or fineness should not be confused with the physical properties of flakes. It is known that appearance is not only dependant on flake size but also on the orientation of flakes, other pigments contained in the coatings, and the illumination and viewing geometry (Kirchner and van den Kieboom *et al.*, 2007) (also see Section 2.5).

Visual texture has also been studied in terms of the modelling of small scale of light interactions within the structural properties of coatings for rendering the appearance (Đurikovič, 2003; Ershov and Kolchin *et al.*, 2001). Although the methods can be used for the simulation of appearance, for example on a computer display, the modelling of the light reflection based on the microstructural properties is computationally expensive. In practice, computationally inexpensive methods are always preferred. Also, there is a need for the automobile refinishing industry to measure the visual texture without knowing the composition of the coatings. Therefore, the approach of the present study is to characterise the visual texture, coarseness and glint, of metallic coatings without any structural information, with less time-consuming and computationally inexpensive methods.

2.6 Texture Analysis

The following sections discuss texture in general and do not focus on the texture of metallic coatings which has been discussed in Section 2.5.1.2. Texture is a term that refers to properties that represent the surface of an object. ASTM (2001) defined texture as; the visible surface structure depending on the size and organization of small constituent parts of a material; typically, surface structure of a woven fabric. Moreover, it was suggested by Pointer (2003) that *physical texture* and *optical texture* be differentiated. Physical texture can be associated with physical, topological, variability in a surface and optical texture is texture associated with spatial variation in appearance caused by non-uniformity of colorant. Texture is a widely used term and perhaps

intuitively obvious but there are, as yet, no precise methods to describe or measure. It is often described subjectively using terms such as coarse, fine, smooth, granulated and rippled.

In computer vision applications, various texture analysis methods have been proposed to compute features of texture based on digital images captured from any surface (see the texture analysis reviews, *e.g.*, in (Chen and Pau *et al.*, 1999)). Texture is represented in a digital image by variation in the pixel values. Computed features are extensively used for any of three purposes: *texture classification*, *texture segmentation* and *texture synthesis*.

Texture classification refers to the process of grouping images of texture into classes, where each resulting class contains similar patterns according to some similarity criterion as illustrated in Figure 2-25. For example, a particular region in an aerial image may belong to agricultural land, a forest region, or an urban area. Each of these regions has unique texture characteristics.

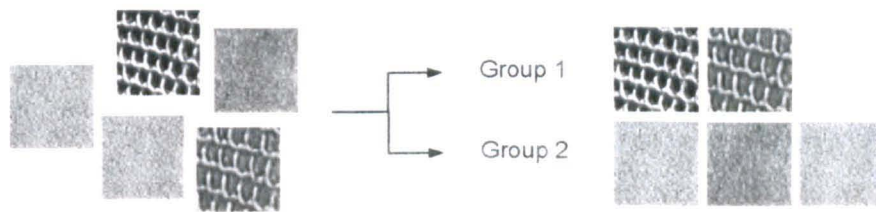


Figure 2-25: Classification of five images contained different characteristics of texture.

Texture segmentation is used to refer to the process of dividing an image into homogeneous regions according to some homogeneity criterion. For example, the four different textures in an image shown in Figure 2-26 (a) can be identified as separate textures as shown in Figure 2-26 (b).

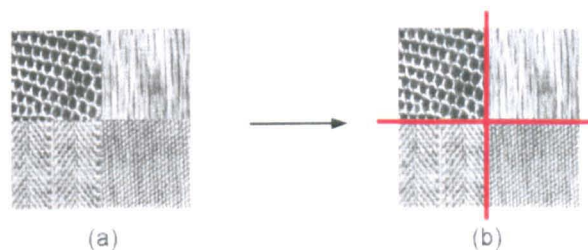


Figure 2-26: (a) An image consisting of four distinct texture regions. (b) The four identified texture regions in the image - the boundary map.

Texture synthesis is often used for image-compression applications. It is also important in computer graphics where the goal is to render object surfaces which are as realistic as possible. An example is given in Figure 2-27.

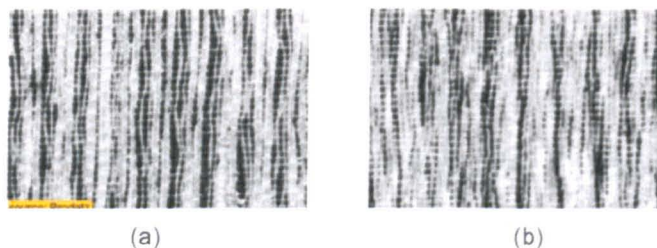


Figure 2-27: (a) An image of the Brodatz textures. (b) The synthesized image based on parameters extracted from the image shown in (a) (Portilla & Simoncelli, 2000).

Although many image texture analysis methods have been proposed, there is no unique method which can be applied because of the wide variation in texture. The following sections review some of the traditional methods in three categories: *statistical methods*, *signal-processing methods* and *structural methods*.

2.6.1 Statistic Methods

Various statistic methods are used to define the texture of an image by a set of statistics extracted from the texture regions. Common descriptive statistics, such as the mean, median, mode and variance can be employed to summarise information according to the image pixel values. Histograms of an image are also often employed to analyse the characteristics of texture by using descriptive statistics. However, histograms do not give any hint as to where the pixels are located within an image (Castleman, 1996). For instance, a black and white checkerboard will have the same histogram as an image in which the top half is black and the bottom half is white.

The two-dimensional spatial properties of an image are obviously important indicators of texture. Therefore, in order to incorporate the spatial information, statistical analyses are often applied not directly to individual pixel values, but to pre-processed information from an image. Some of these methods are introduced in the following sections.

2.6.1.1 Spatial Grey Level Dependence Method

The *spatial grey level dependence method* (SGLDM) (also called the *co-occurrence matrix method*) is concerned with the grey-level spatial independence of pixels and their spatial distribution in a local area (Haralick, 1979; Haralick and Shanmugam *et al.*, 1973). Various statistics describe how the grey levels tend to occur together in pairs (note that this method is usually applied only to a greyscale image).

The SGLDM is based on a matrix $P_{d\theta}(x, y)$ which is the joint occurrence of grey levels x and y for two pixels with a defined spatial relationship. The spatial relationship is defined as a distance d (in terms of the number of pixels) and angle θ . The direction θ usually takes values of either 0° , 45° , 90° and 135° but distance d can be any number of pixels. Figure 2-28 describes the relation of a pair of pixels in an image. For a given distance parameter $d = 1$, both 5 and 1 are paired with the central pixel X for 0° direction (a directional parameter $\theta = 0^\circ$). Similarly, 4 and 8 are paired for $\theta = 45^\circ$, 3 and 7 for $\theta = 90^\circ$ and 2 and 6 for $\theta = 135^\circ$.

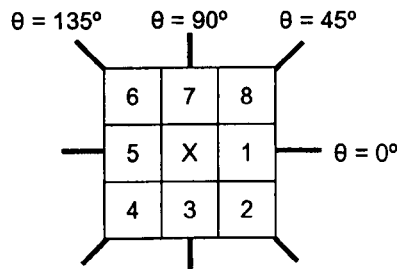


Figure 2-28: Indications of a directional parameter θ for 0° , 45° , 90° and 135° and examples of a pair with a pixel X .

The process of acquisition of a matrix $P_{d\theta}(x, y)$ is best explained by an example using an image I with four grey levels (0-3) as described in Figure 2-29 (a); $I(i, j)$ represents the pixel intensity at a pixel position (i, j) , where i and j represent the rows and columns in the image (i and $j = 1, 2, 3$ and 4 in this example). All the combinations of grey levels x and y for two pixels and a general form for the matrix $P_{d\theta}(x, y)$ (the position of pairs in a matrix $P_{d\theta}(x, y)$) are shown in Figure 2-29 (b). Figure 2-29 (c-f) gives the matrices $P_{d\theta}(x, y)$, which contain the occurrence (the number of times) each of the respective pairs of pixels, for $\theta = 0^\circ$, 45° , 90° and 135° respectively with $d = 1$. Thus, when $\theta = 0^\circ$, the pair of grey levels $(x, y) = (0, 0)$ occurs four times at pixels $(i, j) = (1, 1)$ and $(i, j) = (1, 2)$, pixels $(1, 2)$ and $(1, 1)$, pixels $(2, 1)$ and $(2, 2)$, and pixels $(2, 2)$ and $(2, 1)$.

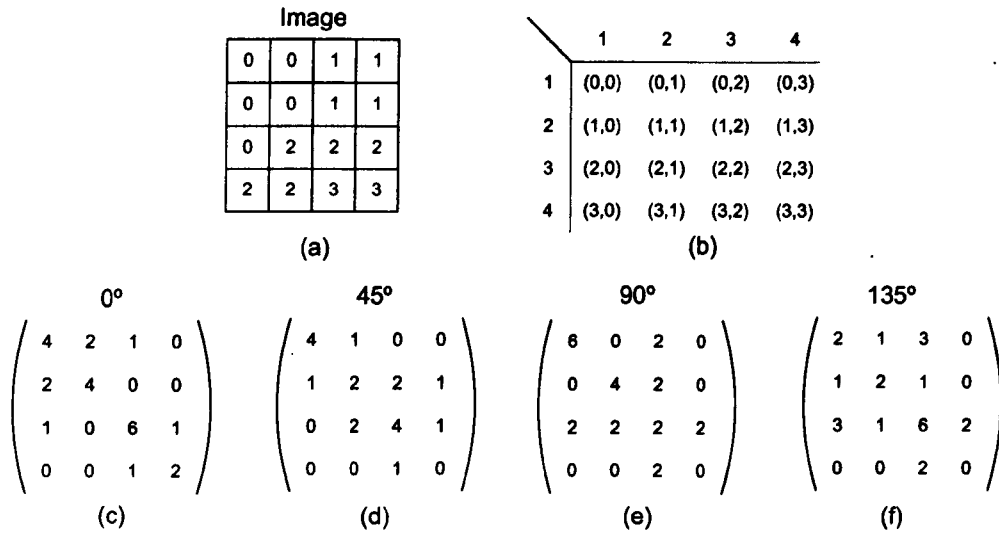


Figure 2-29: (a) A 4×4 image with four grey levels 0-3. (b) General form of and grey level spatial dependence matrix for an image with grey levels 0-3 – the elements stand for the grey levels x and y for two paired pixels. (c)–(f) Matrices $P_{d\theta}(x, y)$ with a distance parameter $d = 1$ for $\theta = 0^\circ, 45^\circ, 90^\circ$ and 135° respectively.

A total of fourteen statistical feature measures based on the SGLDM have been developed to summarise the matrix $P_{d\theta}(x, y)$ numerically and allow meaningful comparisons between various textures (Haralick and Shanmugam *et al.*, 1973). *Energy*, *Contrast* and *Entropy* are commonly selected as texture features. Energy (also called the angular second moment) measures the homogeneity of a texture and is given in Equation 2-16. In a smooth texture, the matrix $P_{d\theta}(x, y)$ will have few entries of large magnitude; therefore the energy will be larger. Contrast (also called inertia) gives an indication of the amount of local variation of intensity presented in an image and can be computed using Equation 2-17. Entropy measures the complexity of a texture which can be calculated using Equation 2-18.

$$\text{Energy (Angular Second Moment)} : \sum_x \sum_y P_{d\theta}^2(x, y) \quad \text{Equation 2-16}$$

$$\text{Contrast} : \sum_x \sum_y (x - y)^2 P_{d\theta}(x, y) \quad \text{Equation 2-17}$$

$$\text{Entropy} : \sum_x \sum_y P_{d\theta}(x, y) \log P_{d\theta}(x, y) \quad \text{Equation 2-18}$$

If the texture is coarse and distance d is small compared with the size of texture elements, the pairs of pixels at the distance d should have similar grey levels. On the other hand, for a fine texture, if distance d is large compared to the size of the texture elements, then the grey levels of the pairs separated by the distance d should usually be

quite different. Therefore, a good way to analyse texture coarseness/smoothness is to apply and compare texture features obtained using various values of distance parameter d (Singh & Singh, 2002). Similarly the directional characteristics of texture can be analysed by varying the direction parameter θ .

Although this is one of the most popular statistical methods and has been successfully used (Connors & Harlow, 1980; Ohanian & Dubes, 1992; Weszka and Dyer *et al.*, 1976), it suffers from a number of difficulties. There is no well-established method of selecting the distance parameter d , and too many texture features are obtained. For images comprising many millions of pixels and quantised to 8 bits (or more), it can be seen that the size of a matrix $P_{d\theta}(x, y)$ is very large and there are also many combinations of the distance parameters. Although the texture features obtained from this method are simple, implementation and comparison are computationally very expensive.

2.6.1.2 Grey Level Difference Method

The *grey level difference method* (GLDM) is similar to the SGLDM, but the GLDM computes texture features based on a vector of absolute difference in grey levels between pairs of pixels in an image. For any given displacement $d = (d_x, d_y)$ where d_x and d_y are integers to indicate the distances in an image, let $I'(i, j) = |I(i, j) - I(i + d_x, j + d_y)|$. A vector $P_d(x)$ is a probability density of $I'(i, j)$. If there are m grey levels, this has the form of an m -dimensional vector. The x th component of $P_d(x)$ is the probability of $I'(i, j)$ having a value of x . Similar to the SGLDM, texture features: *Energy*, *Contrast*, *Entropy* and *Mean* can be computed from the Equation 2-19 to Equation 2-22 (Lew, 2001).

$$\text{Energy} : \sum_x P_d(x)^2 \quad \text{Equation 2-19}$$

$$\text{Contrast} : \sum_x x^2 P_d(x) \quad \text{Equation 2-20}$$

$$\text{Entropy} : - \sum_x P_d(x) \cdot \log(P_d(x)) \quad \text{Equation 2-21}$$

$$\text{Mean} : (1/m) \sum_x x P_d(x) \quad \text{Equation 2-22}$$

If a texture is coarse and the elements of the parameter $d = (d_x, d_y)$ are small compared to the texture element size, the pair of pixels at separation $d = (d_x, d_y)$ usually

has similar grey levels resulting in a small value of $I'(i, j)$. Then, the elements of $P_d(x)$ will be concentrated near $x = 0$.

2.6.1.3 Neighbouring Grey Level Dependence Method

The *neighbouring grey level dependence method* (NGLDM) (Siew and Hodgson *et al.*, 1988; Sun & Wee, 1983) uses an angular independent feature by considering the relationship between a pixel and its all neighbouring pixels at the one time instead of in one direction at a time. This method eliminates angular dependency, while at the same time reducing the computation required to process an image. It is based on the assumption that a grey level dependence matrix of an image can adequately specify the textural information. This matrix is computed from the grey level relationship between every pixel in an image and its all neighbours at the certain distance d . This matrix takes the form of a two-dimensional array $P_{ad}(x, y)$, which can be considered as frequency counts of the grey level variation in an image. The size of $P_{ad}(x, y)$ is $N_g \times N_r$ where N_g is the number of possible grey levels and N_r is the number of possible neighbours to a given pixel in an image. The $P_{ad}(x, y)$ matrix can be obtained by counting the number of pixel pairs having a difference between each pixel $I(i, j)$ and its neighbours (with a certain distance d) equal to certain number a . Figure 2-30 shows an example of an NGLDM matrix P_{ad} computed from a 5×5 image having four grey levels (0-3) with the distance between neighbour $d = 1$ and the difference between its neighbour equal to 0 ($a = 0$).

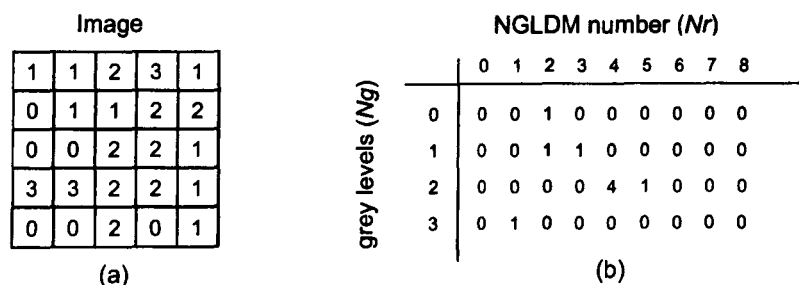


Figure 2-30: An example of a NGLDM. (a) A 5×5 image with four grey levels 0-3. (b) A matrix $P_{ad}(x, y)$ computed from the image in (a).

For the purpose of extracting textural information, the following features: *Small number emphasis*, *Large number emphasis*, *Number non-uniformity*, *Second moment* and *Entropy* can be computed. Small number emphasis measures the fineness of texture in an image and large values of this feature represent finer texture (Equation 2-23).

Large number emphasis measures the coarseness of the texture and large values represent coarser texture (Equation 2-24). Second moment is a measure of the homogeneity (Equation 2-25). In a homogeneous image, there are only a few large numbers of entries in the matrix $P_{ad}(x, y)$. The second moment is the sum of the squares of all entries in the matrix $P_{ad}(x, y)$. Number non-uniformity and entropy are also both related to the coarseness of an image texture (Equation 2-26 and Equation 2-27).

$$\text{Small Number Emphasis : } \sum_x \sum_y \frac{P_{ad}(x, y)}{y^2} / \sum_x \sum_y P_{ad}(x, y) \quad \text{Equation 2-23}$$

$$\text{Large Number Emphasis : } \sum_x \sum_y y^2 P_{ad}(x, y) / \sum_x \sum_y P_{ad}(x, y) \quad \text{Equation 2-24}$$

$$\text{Second Moment : } \sum_x \sum_y P_{ad}(x, y)^2 / \sum_x \sum_y P_{ad}(x, y) \quad \text{Equation 2-25}$$

$$\text{Number Non-uniformity : } \sum_y [\sum_x P_{ad}(x, y)]^2 / \sum_x \sum_y P_{ad}(x, y) \quad \text{Equation 2-26}$$

$$\text{Entropy : } \sum_x \sum_y P_{ad}(x, y) \log(P_{ad}(x, y)) / \sum_x \sum_y P_{ad}(x, y) \quad \text{Equation 2-27}$$

2.6.2 Signal Processing Methods

A common denominator for most signal processing methods is to submit a textured image to a linear or non-linear filter, for example, a *Law filter* (Laws, 1980), a *Laplacian operator* (Lloyd, 2007), a *Fourier domain filtering* and a *Gabor filter* (Kruizinga and Petkov *et al.*, 1999) followed by some statistical analysis. Hence, this approach is sometimes referred to as filtering methods (Randen & Husøy, 1999). Psychophysical research indicates that neural processing in the human visual system analyses an image by decomposing it into the frequency and orientation components (Campbell & Robson, 1968). This is the main motivation for the most signal processing methods that extend feature extraction into the spatial frequency domain. Particularly, many researchers believe that the human visual system acts as a crude Fourier analyser (Hendee & Wells, 1997). As demonstrated by the contrast-sensitivity function (CSF) (see Section 2.2.3), the human visual system processes information within a restricted range of spatial frequencies.

2.6.2.1 Fourier Domain Filtering

Fourier transform has been used as an efficient computational tool in a wide range of applications. A two-dimensional discrete Fourier transform (DFT) converts a digital image into a two-dimensional complex function of energy, also referred to as magnitude, and phase, in the frequency domain (Gonzalez & Woods, 1992). The two-dimensional DFT of an image $I(i, j)$ having size $M \times N$ is defined by Equation 2-28.

$$F(u, v) = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I(i, j) \exp[-p2\pi(ui/M + vj/N)] \quad \text{Equation 2-28}$$

where $p = \sqrt{-1}$, u and v are the frequencies corresponding to i and j respectively, and $0 \leq u \leq M-1$ and $0 \leq v \leq N-1$. The inverse of the function $F(u, v)$ can be written by Equation 2-29.

$$I(i, j) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v) \exp[p2\pi(ui/M + vj/N)] \quad \text{Equation 2-29}$$

where $0 \leq i \leq M-1$ and $0 \leq j \leq N-1$.

It is clear that $F(u, v)$ is a complex function. If the real and imaginary components in $F(u, v)$ are denoted as $F_r(u, v)$ and $F_i(u, v)$, then the energy $E(u, v)$ and phase $\phi(u, v)$ can be calculated by Equation 2-30.

$$\begin{aligned} E(u, v) &= |F(u, v)| = \sqrt{F_r^2(u, v) + F_i^2(u, v)} \\ \phi(u, v) &= \tan^{-1}[F_i(u, v) / F_r(u, v)] \end{aligned} \quad \text{Equation 2-30}$$

The power spectrum $P(u, v)$ of an image can be derived from the Fourier transform function as given in Equation 2-31. It plays important role in image analysis for displaying and analysing the intensity of the image.

$$P(u, v) = |F(u, v)|^2 \quad \text{Equation 2-31}$$

One way to analyse the power spectrum is to divide it into rings (for frequency content) and wedges (for orientation content) as shown in Figure 2-31. The frequency domain is thus divided into regions and then various statistical techniques, such as the total energy in each of these regions, are often applied to describe the information contained and extract the texture feature content.

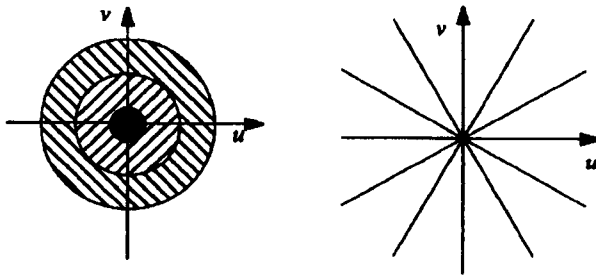


Figure 2-31: The u, v dimensions of the Fourier transform may be split into concentric rings (left) and wedges (right) to encode spatial frequency and orientation information respectively.

2.6.3 Structural Methods

Structural methods assume that a texture image can be regarded as being generated from a set of texture primitives (or texture elements) using a set of placement rules. This approach is better suited to texture having large texture primitives or regular texture. When an image or region which has a simple texture is corrupted by noise, the texture primitive can be first extracted using, *e.g.*, morphological filters, and then their placement rule can be determined. The methods are represented by the *grey level run length method* (Galloway, 1975).

2.6.3.1 Grey Level Run Length Method

A large number of neighbouring pixels of the same grey level represent a coarse texture, a small number of neighbouring pixels represent a fine texture and the lengths of texture primitives in different directions can serve as a texture description. The primitive is a maximum contiguous set of constant grey level pixels located in a line; these can then be described by their grey level, length and direction. Therefore, this method is sometimes called primitive length. The texture features can be based on computation of the continuous probability of the length and grey level of each primitive in the texture. Let $P_{\theta}(x, y)$ be the number of times that an image contains a run of length y consisting of points having grey level x in the given direction θ . The following example in Figure 2-32 shows a 4×4 image having four grey levels (0-3) and the resulting grey level run length matrices for the four principle directions ($\theta = 0^{\circ}, 45^{\circ}, 90^{\circ}$ or 135°).

Image			
0	1	2	3
0	2	3	3
2	1	1	1
3	0	3	0

(a)

		run length																		
		0°				45°				90°				135°						
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4			
grey level	0	4	0	0	0	0	4	0	0	0	0	2	1	0	0	0	4	0	0	0
	1	1	0	1	0	1	4	0	0	0	1	4	0	0	0	1	4	0	0	0
	2	3	0	0	0	2	0	0	1	0	2	3	0	0	0	2	3	0	0	0
	3	3	1	0	0	3	3	1	0	0	3	3	1	0	0	3	5	0	0	0

(b)

Figure 2-32: An example of the Run Length Method. (a) A 4×4 image with four grey levels 0-3. (b) Run length matrices $P_\theta(x, y)$ for each direction $\theta = 0^\circ, 45^\circ, 90^\circ$ and 135° respectively.

To obtain texture features from the matrices, *Short run emphasis*, *Long run emphasis*, *Grey level non-uniformity*, *Run length non-uniformity* and *Run percentage* can be computed. Short run emphasis tends to emphasise short runs and therefore gives greater weight to short runs of any grey level (Equation 2-32). Long run emphasis gives greater weight to long runs of any grey level (Equation 2-33). When runs are equally distributed throughout the grey levels, grey level non-uniformity takes on lower values (Equation 2-34). If the runs are equally distributed throughout the lengths, run length non-uniformity has a lower value (Equation 2-35). Run percentage is the ratio of the total number of runs to the total number of possible runs (Equation 2-36). If all runs had a length of one, it should have its lowest value.

$$\text{Short Run Emphasis : } \frac{\sum_x \sum_y \frac{P_\theta(x, y)}{y^2}}{\sum_x \sum_y P_\theta(x, y)} \quad \text{Equation 2-32}$$

$$\text{Long Run Emphasis : } \frac{\sum_x \sum_y y^2 P_\theta(x, y)}{\sum_x \sum_y P_\theta(x, y)} \quad \text{Equation 2-33}$$

$$\text{Grey Level Non-uniformity : } \frac{\sum_x (\sum_y P_\theta(x, y))^2}{\sum_x \sum_y P_\theta(x, y)} \quad \text{Equation 2-34}$$

$$\text{Run Length Non-uniformity : } \frac{\sum_y (\sum_x P_\theta(x, y))^2}{\sum_x \sum_y P_\theta(x, y)} \quad \text{Equation 2-35}$$

$$\text{Run Percentage : } \frac{\sum_x \sum_y P_\theta(x, y)}{\sum_x \sum_y y P_\theta(x, y)} \quad \text{Equation 2-36}$$

2.7 High Dynamic-Range Imaging

Real-world scenes have a significantly wide range of brightness variation. For instance, luminance levels under the sun at noon may be 100 million (10^8) times higher than under the starlight (Wandell 1995). The human visual system is capable of adapting to these varying conditions and can function over a range of about five orders of magnitude simultaneously within a scene (Reinhard and Ward *et al.*, 2006). In contrast, conventional digital imaging technologies provide only about 8 to 14 bits of information for each of red, green and blue channel of each pixel. Having only 8 to 14 bits per channel is not enough to represent many real-world scenes. Therefore, a set of techniques called *high dynamic-range imaging* (HDR imaging) has been developed which aim to increase the limited range of levels that conventional technology can provide.

HDR images were originally adopted for the production of more realistic images for computer games and other computer applications. Later, the interest in HDR imaging extended to the idea of producing better images of natural scenes from images captured using digital cameras where limitations in sensor design prevent the capture of the high dynamic range of typical scenes. In contrast to HDR, the dynamic range which conventional cameras can capture is called *low dynamic-range* (LDR). As a result, when capturing a scene which contains a higher dynamic range than a conventional camera can capture, the image produced will either be too dark in some areas, and possibly saturated, or even both. To cover the full dynamic range of such a scene, a series of images is captured using multiple exposures, a range of camera apertures, or by using a set of neutral density filters over the lens of the camera. However, the method of using various apertures is not recommended due to the effect of aperture on the depth of a field (Robertson and Borman *et al.*, 2003). The most frequently used method is to use multiple exposures. Clearly, an over-exposed image can be saturated in light areas, but contains good detail in dark areas. On the other hand, an under-exposed image exhibits less saturation in light areas but can lose detail in the shadows. The complementary nature of these images allows combining them into a single HDR image. Various algorithms have been introduced in literature (Debevec & Malik, 1997; Krawczyk and Goesele *et al.*, 2005; Mitsunaga & Nayar, 1999; Nayar & Mitsunaga, 2000; Robertson and Borman *et al.*, 2003; Xiao and DiCarlo *et al.*, 2002). An example is given in Figure 2-33. The images on the left side are LDR images which have been captured using a range of exposures. A single HDR image derived from the LDR

images is shown on the right. Since the HDR image could not be displayed on a conventional 8 bit display, tone mapping needs to be applied, which converts HDR images into a viewable format on LDR devices such as cathode-ray tube (CRT) displays, liquid-crystal displays (LCD) and printers (Fattal and Lischinski *et al.*, 2002; Kuang and Yamaguchi *et al.*, 2004; Larson and Rushmeier *et al.*, 1997; Meylan, 2006; Meylan & Süssstrunk, 2006; O'Malley, 2006).

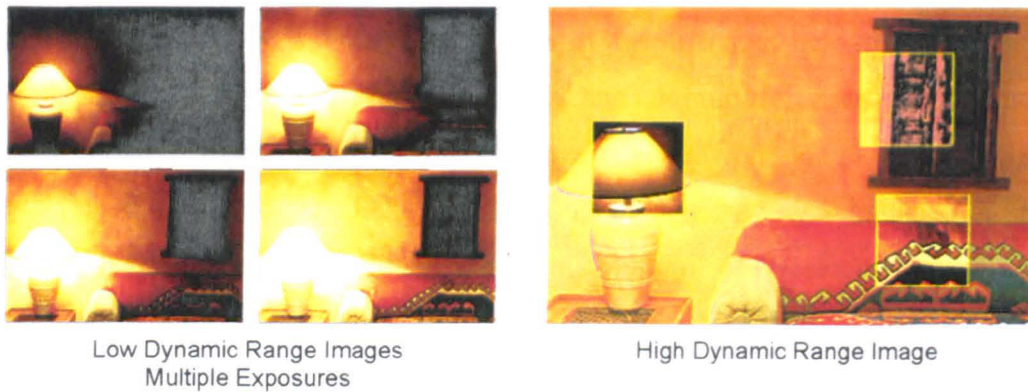


Figure 2-33: An HDR image (right) created from four LDR images (left) (CAVE, 2007).

2.8 Image Acquisition and Display Devices

The following sections review image acquisition and display device technologies. The basic structures of a digital camera and a liquid-crystal display (LCD) are first reviewed, followed by an overview of methods for device characterisation.

2.8.1 CCD Camera Systems

There are two main types of sensors used for current digital camera systems, which are a *charge-coupled device* (CCD) sensor and a *complimentary metal-oxide semiconductor* (CMOS) sensor. Currently, CCD sensors tend to have less noise, high light sensitivity, less pixel cross-talk than CMOS sensors, but CMOS sensors consume less power and inexpensive compared to CCD sensors (Litwiller, 2005). CCD sensors have been mass-produced for a longer period of time, so they are more mature and tend to create higher quality of images.

For a digital camera incorporated with a CCD sensor, three main components are a lens, a CCD sensor and a digital signal processing chip (Lee, 2005). The amount of

light that reaches the sensor can be controlled by both the aperture size and the shutter speed (the exposure time). The sensor has spatially discrete sensing elements (known as pixels). When photons strike a sensor, electronic charge is produced whose magnitude is proportional to the intensity of the incoming radiation during the exposure time. The signal at each pixel is then digitised by an analogue-to-digital converter. Signal processing then produces a digital image in the required format. This processing takes a variety of forms including interpolation for missing pixels, tone scale correction and digital compression (Lee, 2005).

There are two different basic types of CCD-based digital camera (Ippolito, 2002). One type utilises three CCD sensors to capture the image. This type of camera receives the light with three different filters entirely covering each individual sensor. This is similar to the human visual system (see Section 2.2.2), which is sensing different spectral bands with three types of cones, *L*-, *M*- and *S*-cones. It has been argued that if the sensitivities of the colour filters combined with that of the sensor are linear transforms of the LMS sensitivities or the CIE CMFs (*Luther condition* (Ohta, 1982)), there will be no difference between what the camera system 'sees' and what a human observers sees (Lee, 2005). However, in practice this is not the case: the Luther condition is not realisable due to the difficulty in manufacturing such filters. If the filters were to mimic the human cone spectral sensitivity functions, the system signal-to-noise performance may be unacceptable. Since the human *L*-, *M*- and *S*-cones have a large overlap in their sensitivities, it is difficult to create signals without overly magnifying the noise (Lee, 2005). Therefore, filters responding in approximately the red, green and blue regions of the visible spectrum are usually utilised. Each sensor responds to the light intensity for one colour (red, green or blue) with a full sensor resolution simultaneously, and the camera must combine the output of the three CCD sensors to create a single image. This produces a high quality image, but, having three CCD sensors significantly increases the price of the camera.

An alternative to using three CCD sensors is a single CCD sensor as used in most consumer cameras. The individual pixels in the sensor are coated with one of three filters such that each pixel senses the incoming light in only one of the three specular bands. One of the most popular colour filter arrays is called Bayer pattern (Bayer, 1976) which uses the filters responding in the red, green and blue regions of the visible spectrum, and uses twice as many green elements as red or blue as illustrated in Figure 2-34. This colour filter array is designed to give approximate luminance sensitivity with

the green filter because the human visual system is more sensitive to the luminance component than the chromatic components (see Sections 2.2.2 and 2.2.3). The increased number of green sensitive pixels improves the spatial sampling of the luminance signal. Since each pixel is filtered to record only one of three colours, two thirds of the colour information is missing. In order to obtain a full colour image, a ‘demaicing’ algorithms has to be applied to interpolate between the information from adjacent pixels.

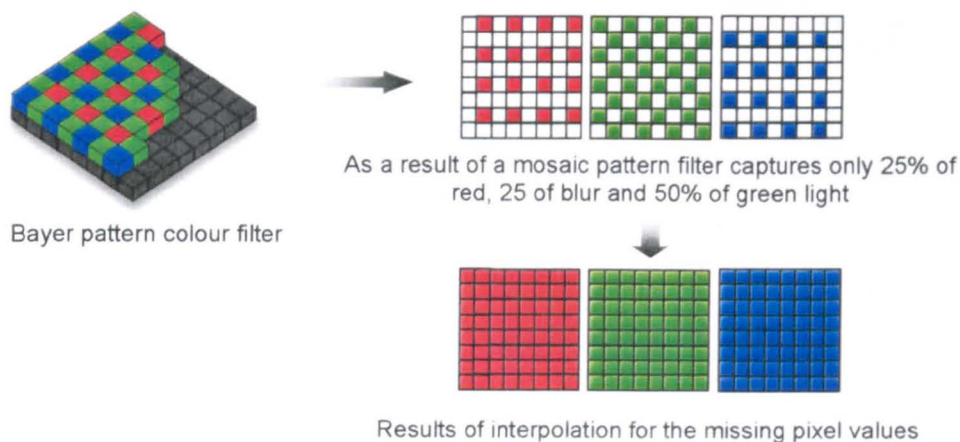


Figure 2-34: Schematic diagram of Bayer pattern and its process.

2.8.2 Display Systems

Liquid-crystal displays (LCDs) are now replacing cathode-ray tube (CRT) displays. In comparison with CRT displays, LCDs require lower voltage and smaller power sources, and a distinct advantage is their lower physical size and weight.

Molecules forming liquid-crystals are a main component of LCDs. Liquid-crystals are substances that share properties of both liquid and crystalline states. Molecules in a typical fluid, for example water, have random locations and are rotated in random directions. Conversely, molecules in a molecular crystal have ordered positions on a lattice and point in well-defined directions. The essential properties of liquid-crystals are their optical and electromagnetic anisotropy and they have an orientation order where the long axis of the molecules tends to align in a preferred direction (Lee, 2005). When elongated molecules are placed in an electric field, they are polarized. The molecules rotate to rearrange their orientation such that the dipole moments are aligned with the external field and this induced an orientation change and also changes the optical transmission of the material. This is the basic phenomenon which enables liquid-crystal devices to work as electronic displays. Depending on the

type of orientation order, there are three different anisotropic phases in liquid-crystals: *nematic*, *smectic* and *cholesteric* (Lee, 2005). For display applications, nematic and twisted nematic (TN) phases are commonly used. The nematic phase is characterised by the long-range orientation order. Therefore, long molecular axes align roughly in their preferred orientations, but their relative positions are random. Figure 2-35 gives a schematic diagram of the physical structure of a typical LCD (TN liquid crystal and active matrix (LG Philips, 2007)). An LCD consists of a TN liquid crystal sandwiched between two plates of glass (HORIBA Jobin Yvon, 2007). The glass plates are coated with a transparent conducting material (made of indium tin oxide (ITO)) which acts as an electrode. On one side of the glass, there are a common electrode and colour filters which generate the colour (red, green and blue) information for colour displays. On the other side of the glass is a pixel electrode which can be patterned to form rows and columns of a *passive matrix* display or the individual pixels of an *active matrix* display. Therefore, in the case of a passive matrix display, the pixels are addressed one at a time by row and column addresses. Thus, an effective voltage applied to liquid crystals must average the signal voltage pulses over several frame times, which results in a slow response time and a reduction of the maximum contrast ratio. For an active matrix, the switching devices are located at each sub-pixel to control each pixel separately. A pixel, the smallest unit of a picture image, is formed by three sub-pixels with red, green, and blue outputs. Therefore, active-matrix displays look brighter and sharper than passive-matrix displays of the same size, generally have a quicker response time and produce much better images. Most transistors are made of deposited thin films, which are therefore called thin-film transistors (TFTs). The voltage between the pixel electrode and the common electrode which is supplied by a backlight, creates the electrical field that is applied to the liquid-crystal molecules which determine the amount of movement of the liquid-crystals required to change the light-transmitting properties. The red, green and blue colour filters modify the output spectra to produce the desired colour stimulus at a spatial location on the image (Lee, 2005).

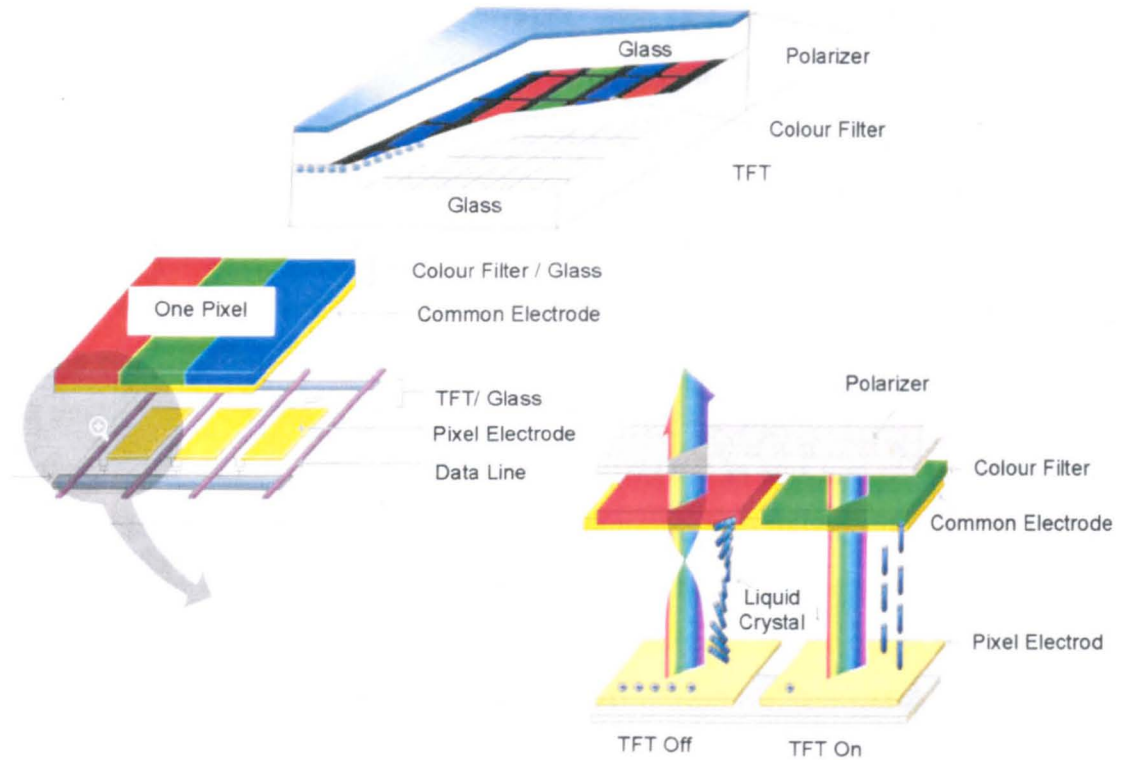


Figure 2-35: Schematic diagram of the physical structure of a typical LCD (TN liquid crystal and active matrix) (adapted from (LG Philips, 2007)).

2.8.3 Device Characterisation

A digital camera and an LCD quantify colours corresponding to the amount of light passing through their colour filters as described in Sections 2.8.1 and 2.8.2. These devices often use three filters which respond broadly in the red, green and blue regions of visible spectrum. Therefore, their input or output values are often called *RGB values*, although the spectral sensitivity curves of the filters are different depending on cameras and LCDs. Hence, for example, the output RGB values of identical scenes captured by two cameras can be different depending on the filters used, Figure 2-36. Similarly, identical input RGB values to two LCDs can result in different perceived colours. RGB values describing colours are therefore called *device-dependent* values and the associated colour space is termed a device-dependent colour space. On the other hand, *device-independent* values or colour spaces express colours are not subject to the devices; an example being the CIE systems and associated colour spaces (Johnson, 1996).



Figure 2-36: An example of device-dependent values. Different cameras give different RGB values from the same scene.

Calibration and *characterisation* are two important procedures that must be applied to capture and display devices to be used for scientific research. Calibration refers to achieving a predefined state for a device system and is an essential first step to obtain repeatable data. For digital cameras, the settings involve lighting conditions, exposure time, aperture settings, *etc.* For displays, calibration includes the settings of the system white point, the system gamma, *etc.* Techniques for the calibration of cameras and displays have, for example, been described by Cheung (2004) and Berns (1996) respectively.

Characterisation defines the relationship between device-dependent and device-independent colour space (Johnson, 1996). A *characterisation model* is a tool to transform any colour from one space to another. Many characterisation models have been proposed but no one type of model gives optimal results to every type of device. Green (2002b) classified the basic models into the following three groups according to the approach to generating the models: *physical models*, *numerical models* and *look-up tables*. Physical models include terms for various physical properties of the device, for example as the absorbance, scattering and reflectance of colorants and substrates, or the gain, offset and gamma of display devices. The *GOG (Gain-Offset-Gamma) model* (Berns, 1996) is an example of a physical model for the characterisation of a display device. Similarly, the *Kubelka-Munk* and *Yule-Nielsen* models are physical models which can be used as the basis of a characterisation model for a printer (Green, 2002a). Numerical models define a series of coefficients, usually from a set of known samples, with no prior assumptions about the physical behaviour of the devices and associated media. Examples of numerical methods include *linear* and *polynomial regression models* and *artificial neural networks* (Cheung, 2004). Look-up tables define the conversion between a device colour space and a CIE colour space as a series of coordinates (a table of data) within the colour space, and interpolate the values for intermediate coordinates. The entries in the look-up table can be determined either by

direct measurements or through a physical or numerical model. In practice, two or more of these approaches can be combined to provide the necessary transformation. The details of some of these models are given in the following sections. Prior to this, performance of the experimental practices of previous studies are reviewed.

Linear and polynomial regression models, with least-squares fitting, have been applied to digital camera and scanner characterisation and satisfactory performance has been obtained. Kang (1997) applied these models to a flat-platen Sharp JX450[®] scanner comparing different numbers of polynomial terms (3×3 , 3×6 , 3×8 , 3×9 , 3×11 and 3×20) and various data sets (different numbers of *training data* which are samples used to derive a transfer matrix). The details of polynomial modelling will be discussed in Section 2.8.3.1. The most accurate performance obtained, in terms of the mean ΔE^*_{ab} of *test data* which are samples used to test the performance of the derived model, was a value of 1.50 using a 3×20 -term polynomial model with a Kodak Q60 photographic standard chart (105 colour patches in the standard as a training data set and 131 as a test data set). Hong and Luo *et al.* (2001) applied polynomial models to digital camera characterisation. In their study, an Agfa digital StudioCam camera and the colour patches in an ANSI IT8.7/2 chart as data were used. The mean value of $\Delta E_{CMC(1:1)}$ was 1.21 using a 3×11 -term polynomial model with 96 and 168 colour patches as a training and test data sets respectively. Similarly, Cheung (2004) carried out a comparison study of characterisation models with different numbers of polynomial terms (3×3 , 3×4 , 3×5 , 3×10 , 3×20 and 3×35) using an Agfa digital StudioCam camera and colour patches in a GretagMacbeth ColorChecker DC chart as a training data set and selected 50 colour patches from the Natural Color System (NCS) as test data. The model with 3×20 terms gave the most accurate result with a median ΔE^*_{ab} of 1.13 (a mean ΔE^*_{ab} of 1.28). Also, neural network models were evaluated in the same manner and a similar accuracy found with a median ΔE^*_{ab} of 1.24 (a mean ΔE^*_{ab} of 1.44). Although the polynomial regression and neural network models showed similar performance, it was concluded that there was no advantage in using the neural network model over the polynomial model, since the neural network model is more computationally complex and time consuming.

The GOG model (see Section 2.8.3.2) is a well-established characterisation model that can be applied to a CRT display. The model accuracy, in terms of ΔE^*_{ab} , of less than 1.0 has been reported from various studies (Berns, 1996; Cui and Luo *et al.*, 2001; Gibson & Fairchild, 2000). Compared with CRT technology, the characteristics of

LCDs have not matured. Therefore, if the optoelectronic transfer function (OETF) of an LCD is similar to that of a typical CRT display, the GOG model can be applied for characterisation. However, the OETFs of LCDs are not always similar to those of CRTs and is consequently the reason why various different analytical models have been recently introduced (Gibson & Fairchild, 2000; Kwak & MacDonald, 2001; Sharma, 2002; Yoshida & Yamamoto, 2001). A model proposed by Day and Taplin *et al.* (2004) shares similarities with the GOG model, but is different in the method for charactering the OETF. Values of ΔE_{00} were obtained which ranged from 0.1 to 0.4. *Three-dimensional (3D) look-up table with interpolation* can be used to characterise displays as well as other devices such as printers and scanners (Hung, 1993; Kang, 1997). This method provides high accuracy, especially for devices which have a complex nature, such as plasma display panels (PDP) or when analytical models do not work in practical systems. However, the disadvantage is that it requires a large number of samples and is thus considered time consuming.

In the following sections, the structure of linear and polynomial models is first introduced. The GOG model and the analytical model proposed by Day and Taplin *et al.* (2004) for characterising a display are then described.

2.8.3.1 Linear and Polynomial Regression Model

The linear and polynomial regression models assume that the correlation between colour spaces can be approximated by a set of simultaneous equations. These can be express using a formula as shown in Equation 2-37. The example of a linear transform of XYZ values from RGB values is given in Equation 2-38.

$$I = D \times M \quad \text{or} \quad D = I \times M^{-1} \quad \text{Equation 2-37}$$

$$[X \ Y \ Z] = [R \ G \ B] \begin{bmatrix} a_{1,1} & a_{2,1} & a_{3,1} \\ a_{1,2} & a_{2,2} & a_{3,2} \\ a_{1,3} & a_{2,3} & a_{3,3} \end{bmatrix} \quad \text{Equation 2-38}$$

where I represents the device-independent values, D is the device dependent values, M is the transfer matrix and $a_{1,1}$ to $a_{3,3}$ are the elements of the transfer matrix that links these two colour spaces. In order to derive a transfer matrix, a set of data consisting of device-dependent values and device-independent values of uniform colour samples are required. In the case of cameras and displays, a set of RGB values and the CIE XYZ

tristimulus values can be used. The most efficient method to obtain a data set for input devices (*e.g.*, digital cameras or scanners) is to capture a colour chart containing a wide range of colour patches of known XYZ values (Westland & Ripamonti, 2004). A Macbeth ColorChecker DC chart or an ANSI IT8.7/1 or 2 are often used, since they include colour patches reasonably uniformly distributed in colour space. For output devices (*e.g.*, displays and printers), output of created sample colour patches of a range of device values is measured by any appropriate colour measurement instrument.

In performing regression, a transform matrix M needs to be derived from a set of known XYZ and RGB data set which is described in Equation 2-39.

$$M_{3 \times 3} = D_{n \times 3}^{-1} I_{n \times 3} \quad \text{Equation 2-39}$$

where subscripts n and the numbers denote the matrix size. They represent the numbers of data (n RGB values and n XYZ values). If there are three suitable samples available, then $n = 3$ and M is easy to compute from the inverse of D and has a single and unique solution. However, if $n > 3$ then there is no exact solution (an over-determined system), since D has become a non-square matrix. This is always the case for device characterisation, since the number of colour samples that are used to define the transfer matrix is almost always greater than the number of coefficients. Therefore, a least-squares method must be used to determine the best-fit approximation of the coefficients in the transfer matrix by minimising the error matrix E between I and DM given in Equation 2-40.

$$\|E\|_F^2 = \|I - DM\|_F^2 \quad \text{Equation 2-40}$$

The transfer matrix M can be solved in terms of I and D written in Equation 2-41.

$$M = (D^T D)^{-1} D^T I \quad \text{Equation 2-41}$$

When the linear model does not yield a sufficiently accurate transformation, a polynomial regression can be used. An example of a second order polynomial model is given in Equation 2-42.

$$[X \ Y \ Z] = [R \ G \ B \ R^2 \ G^2 \ B^2] \begin{bmatrix} a_{1,1} & a_{2,1} & a_{3,1} \\ a_{1,2} & a_{2,2} & a_{3,2} \\ a_{1,3} & a_{2,3} & a_{3,3} \\ a_{1,4} & a_{2,4} & a_{3,4} \\ a_{1,5} & a_{2,5} & a_{3,5} \\ a_{1,6} & a_{2,6} & a_{3,6} \end{bmatrix} \quad \text{Equation 2-42}$$

The coefficients $a_{1,1}$ to $a_{3,6}$ are found in a similar manner as in Equation 2-39 to Equation 2-41. As a higher order polynomial model is used, the size of the matrix D and M are adjusted accordingly. However, it is desirable to keep the polynomial order as low as possible (Green, 2002b).

2.8.3.2 GOG Model

Berns *et al.* (Berns, 1996; Berns and Grozynski *et al.*, 1993; Berns and Motta *et al.*, 1993) proposed the GOG (Gain-Offset-Gamma) model based on their investigation of the physical properties of CRT displays. Unlike LCDs, the technology for CRT displays is more stable and the GOG model has been widely used and adopted in colour management software (Kato and Deguchi *et al.*, 2001b).

The GOG model consists of two stages. The first stage models the optoelectronic transfer function (OETF) of a CRT display, in which a non-linear relationship between the digital counts d and the spectral radiance L_λ as described in Equation 2-43 for the red channel as an example.

$$L_{\lambda r} = \begin{cases} k_{\lambda r} \left[a_r \left[(v_{\max} - v_{\min}) \left[\frac{LUT_r(d_r)}{2^N - 1} \right] + v_{\min} \right] + b_r - v_{cr} \right]^{\gamma_r} \\ \quad \text{if } v_{cr} \leq a_r \left[(v_{\max} - v_{\min}) \left[\frac{LUT_r(d_r)}{2^N - 1} \right] + v_{\min} \right] + b_r \\ 0 \quad \text{if } v_{cr} > a_r \left[(v_{\max} - v_{\min}) \left[\frac{LUT_r(d_r)}{2^N - 1} \right] + v_{\min} \right] + b_r \end{cases} \quad \text{Equation 2-43}$$

where the subscript r refers to the red channel, LUT represents the video look-up table, N is the number of bits in the digital-analogue converter (DAC), v_{\min} and v_{\max} are minimum and maximum voltages of the video-signal generator, a_r and b_r are the CRT video amplifier gain and offset, v_{cr} is the cut-off voltage that defines zero beam current, γ_r is the gamma of the channel and $k_{\lambda r}$ is a spectral constant to account for the particular CRT phosphors and faceplate combination. Generally, an accurate physical model of display behaviour is not used for characterisation purposes, rather the relationship between the digital counts d and the spectral radiance $L_{\lambda r}$ is adopted as express in Equation 2-44 (Westland & Ripamonti, 2004).

$$L_{\lambda_r} = \begin{cases} L_{\lambda_r, \max} \left(g_r \frac{d_r}{2^N - 1} + o_r \right)^{\gamma_r} & \text{if } \left(g_r \frac{d_r}{2^N - 1} + o_r \right) \geq 0 \\ 0 & \text{if } \left(g_r \frac{d_r}{2^N - 1} + o_r \right) < 0 \end{cases} \quad \text{Equation 2-44}$$

It can be useful to think of the coefficient g_r and o_r as the system gain and offset respectively. The implication of Equation 2-44 is that although the CRT has an inherent fixed gamma, the effective gamma of the system can be set according to the offset and gain controls. Under the assumption that the spectral radiance L_{λ} , at any level of phosphor excitation, can be related to the maximum spectral radiance (Katoh and Deguchi *et al.*, 2001b), a radiometric scalar s , the ratio of a given signal to the maximum signal, throughout the wavelengths can be applied instead. This property is often referred to as *chromaticity constancy of primaries* (Gibson & Fairchild, 2000) or *channel scalability* (Day and Taplin *et al.*, 2004). In practice, a radiometric scalar s can be expressed as the linearised normalised digital counts d using Equation 2-44. This can be written as Equation 2-45.

$$s_r = \begin{cases} \left(g_r \frac{d_r}{2^N - 1} + o_r \right)^{\gamma_r} & \text{if } \left(g_r \frac{d_r}{2^N - 1} + o_r \right) \geq 0 \\ 0 & \text{if } \left(g_r \frac{d_r}{2^N - 1} + o_r \right) < 0 \end{cases} \quad \text{Equation 2-45}$$

where the normalisation procedure requires that the system gain and offset are equal to unity. Therefore, the offset can be express as $o_r = (1 - g_r)$. Analogous equations can be used for the other two channels.

At the second stage, the scalars can be related to the XYZ values using a simple linear transform as shown in Equation 2-46. A primary matrix consisting of the measured XYZ values of each of the red, green and blue channels at maximum digital count is applied for this transformation by assuming that the *channel additivity rule* (Gibson & Fairchild, 2000) holds. However, in practice, this assumption does not hold in cases where a small amount of luminance can be detected at the black level as a result of measurable flare. There are mainly three types of flare which can be described. These are external flare which comes from reflections on the display from ambient light; internal flare which comes from internal scattering; and a second type of internal flare which comes from output from other channels at the same pixel location (Katoh and Deguchi *et al.*, 2001a; Katoh and Deguchi *et al.*, 2001b). It is characterised by the

black level which is defined as the output XYZ values from zero digital counts for all channels ($d_r = d_g = d_b = 0$). Therefore, Equation 2-46 can be replaced by Equation 2-47.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} X_{r,\max} & X_{g,\max} & X_{b,\max} \\ Y_{r,\max} & Y_{g,\max} & Y_{b,\max} \\ Z_{r,\max} & Z_{g,\max} & Z_{b,\max} \end{bmatrix} \begin{bmatrix} S_r \\ S_g \\ S_b \end{bmatrix} \quad \text{Equation 2-46}$$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} X_{r,\max} - X_k & X_{g,\max} - X_k & X_{b,\max} - X_k \\ Y_{r,\max} - Y_k & Y_{g,\max} - Y_k & Y_{b,\max} - Y_k \\ Z_{r,\max} - Z_k & Z_{g,\max} - Z_k & Z_{b,\max} - Z_k \end{bmatrix} \begin{bmatrix} S_r \\ S_g \\ S_b \end{bmatrix} + \begin{bmatrix} X_k \\ Y_k \\ Z_k \end{bmatrix} \quad \text{Equation 2-47}$$

where X , Y and Z define the tristimulus values, the subscript *max* represents the maximum output and X_k , Y_k and Z_k are the black level (minimum) output. Three model parameters, gain, offset and gamma, for each channel can be estimated by solving for minimum error between estimated scalars for a training data set (usually a grey scale) using the inverse of Equation 2-47 and using Equation 2-45.

2.8.3.3 Day, Taplin and Berns Method

A characterisation model has been proposed for LCD characterisation by Day, Taplin and Berns (Day and Taplin *et al.*, 2004). Similar to the GOG model, there are two stages; one is to characterise OETF and the other is the transformation between a radiometric scalar and the XYZ values. OETFs of a CRT display can be characterised using gain, offset and gamma values based on the physical characteristics, whereas OETFs of an LCD depend on the technology or system of operation used. Therefore, a solution to characterise OETFs is to build one-dimensional look-up tables (LUTs), generally formed by subsampled measurements and linear or non-linear interpolations, as described by Equation 2-48.

$$s_r = LUT(d_r) ; s_g = LUT(d_g) ; s_b = LUT(d_b) \quad \text{Equation 2-48}$$

$$(0 \leq s_r, s_g, s_b \leq 1)$$

where d defines the digital counts, s defines the radiometric scalar and the subscripts r , g and b represent the red, green and blue channels respectively. The scalars are constrained to values between zero to unity by taking the ratio of a given signal to the maximum signal. The second stage is to relate the scalars to the XYZ values using a simple linear transform. As with the GOG model, flare needs to be considered. Especially for LCDs, there is often significant radiant output at the black level caused

by the liquid-crystal having a minimum transmittance factor well above zero (Day and Taplin *et al.*, 2004). Thus, the relationship between the scalars and the XYZ values can be also described using Equation 2-47. The relations described in Equation 2-47 assume perfect chromaticity constancy of primaries and channel additivity (Gibson & Fairchild, 2000). However, it is known that the spectral transmittance of liquid-crystals varies as a function of voltage. The peak wavelength shifts toward shorter wavelengths with decreasing transmittance (Day and Taplin *et al.*, 2004). Therefore, in practice, these assumptions are not met. In order to compensate for this, the coefficients in Equation 2-47 are optimised by minimising the colour differences between measured and estimated colours, sampled over the colour gamut of the LCD. Then, the LUTs are re-computed according to the optimised coefficients. The worked flowchart is given in Figure 2-37.

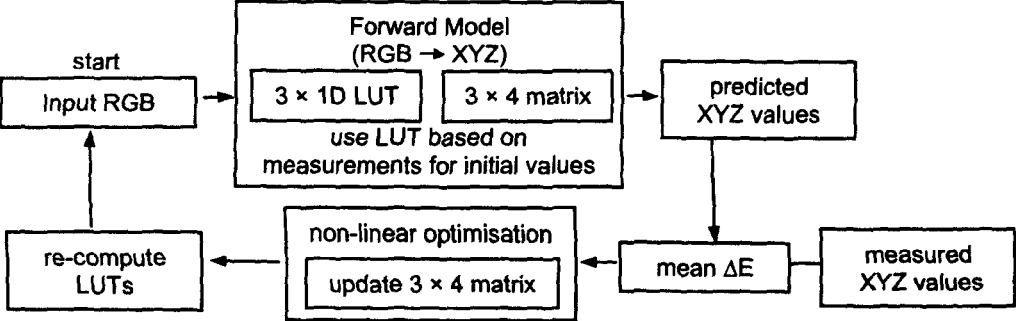


Figure 2-37: A worked flowchart of the Day, Taplin and Berns characterisation model (Day and Taplin *et al.*, 2004).

2.9 Psychophysics: Quantitative Methods for Perceptual Responses

Much of the knowledge about the human visual system has been revealed from the interaction between the study of neural mechanisms and the investigation of perceptual responses using the methods of psychophysics. For example, as has been shown in the previous sections, CIE colorimetry (see Section 2.3) and the human CSF (see Section 2.2.3) have been established based on psychophysical studies. Psychophysics is a scientific discipline designed to determine the relationship between physical measurements of stimuli and sensations or perceptions that those stimuli evoke. Many different experimental techniques can be used to measure the perceptions of stimuli. For visual psychophysics, the experimental methods tend to fall into the two broad classes: *threshold* and *scaling* methods (Johnson & Fairchild, 2003). Threshold

methods include detection, discrimination and matching experiments. The experimental method for measuring the human CSF is an example of a threshold method. Also, the basis of CIE colorimetry was derived using threshold methods; specifically the *(colour) matching method*. The second class, scaling methods, is appropriate to generate the relationship between the physical and the perceptual magnitudes of a stimulus. Scaling is based on the assignment of numbers to physical stimuli (objects) according to stated rules. S.S Stevens identified four basic types of scales: *nominal*, *ordinal*, *interval* and *ratio* (Luce & Suppes, 2002). Nominal scales are relatively trivial. They merely use numbers instead of names or labels to identify the stimulus. Even if numbers are used, they are just labels which mean they do not possess any numerical properties. Ordinal scales are derived by the observer placing the stimuli in an order defined by some attribute: they give no information about the distances between the scale values. In the case of interval scales, the differences between scale values have equal weight in terms of the scaled attribute. However, generally, an origin of the scale is unknown and arbitrary, *i.e.*, there is no meaningful zero point. Ratio scales are an interval scales with a meaningful zero point.

There are many experimental methods for using nominal, ordinal, interval and ratio scales, however, the following sections only introduce the experimental scaling methods used in the present study: the *categorical judgement scaling method* and the *magnitude estimation method*. These methods were applied to provide scales of suprathreshold perceptual difference in terms of perceptual coarseness and perceptual glint of a set of metallic coating panels.

2.9.1 Categorical Judgement Scaling Method

The categorical judgement scaling method is designed to measure the perception of the observer on an equal-appearing interval scale of categories. The general categorical judgement scaling method requires observers to assign stimuli to predefined categories according to the size of the perceptual attribute of interest. These categories are usually specified to observers either as numbers (*e.g.*, 1, 2 and 3) or as names (adjectives are generally used, *e.g.*, good, better or best). The number of categories is usually between three and twenty although usually an odd number of categories from five to eleven are used. If there are too few categories available for observers, it is possible that they are not able to discriminate between stimuli. On the other hand, too

many categories causes errors because observers may be tempted to assign stimuli having the same perception to different categories (Park and Jeon *et al.*, 2004). This is however, dependent on the attributes being scaled. Most observers typically distinguish only about seven different categories so that additional categories may contribute very little useful information (Engeldrum, 2000).

As a result of applying this method, the categories corresponding to the given stimuli are obtained from each observer. Then data analysis is used to derive *scale values* that represent the perceptual responses of the observers to the stimuli.

The simplest way of determining scale values is to take the arithmetic mean of categories for each stimulus. This method is called the *mean-category value method* (Bartleson, 1984). Underlying such a method is the tacit assumption that observers are capable of keeping the intervals between category boundaries psychologically equal, as they assign stimuli to various categories. However, this is often taken as an unwarranted assumption, as this method is influenced by many factors that have a tendency to lead to unequal intervals, for example, stimulus range and stimulus sequence (Gescheider, 1997). Yet it is frequently found that this simple mean-category value method yields scale values that are very close to those determined by a more sophisticated technique known as the *categorical-judgement method* following *Torgerson's law of categorical judgement* (Torgerson, 1967).

Torgerson's law of categorical judgement was proposed to consider the issues raised by the mean-category value method and is an extension of *Thurstone's law of comparative judgements* (Thurstone, 1959). The assumptions made by Torgerson (1967) were;

"1. The psychological continuum of the subject can be divided into a specified number of ordered categories or steps.

2. Owing to various and sundry factors, a given category boundary is not necessarily always located at a particular point on the continuum. Rather, it also projects a normal distribution of positions on the continuum. Again, different category boundaries may have different mean locations and different dispersions.

3. The subject judges a given stimulus to be below a given category boundary wherever the value of the stimulus on the continuum is less than that of the category boundary."

According to the underlying assumption of the Torgerson's law of categorical judgement, Figure 2-38 illustrates an example of the distribution of observers' judgements on the stimulus (full line) and the distributions of category boundaries (dashed line) along a perceptual scale. This can be formulated in terms of the differences between scale values and category boundaries as given in Equation 2-49.

$$T_j - S_i = z_{ij} \sqrt{\sigma_i^2 + \sigma_j^2 - 2\rho_{ij}\sigma_i\sigma_j} \quad (i=1,2,\dots,m)(j=1,2,\dots,n-1) \quad \text{Equation 2-49}$$

where T_j is the mean location of the upper j th category boundary, S_i is the scale value of stimulus i , z_{ij} is the unit normal deviate (z-score) corresponding to the population of times stimulus i sorted below the category boundary j , σ_j is the dispersion (standard deviation) of the j th category boundary, σ_i is the dispersion (standard deviation) of stimulus i , ρ_{ij} is the correlation between the momentary positions of the category boundary j and stimulus i , n is the number of categories; namely there are $n-1$ category boundaries for n categories; m is the number of stimuli.

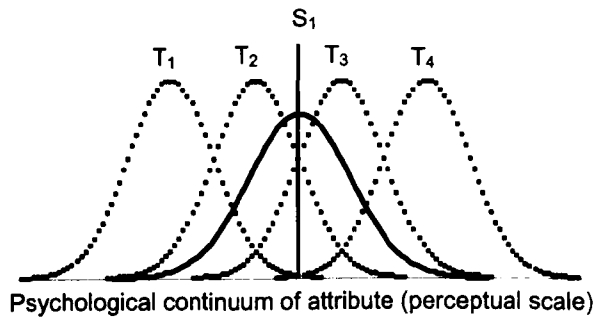


Figure 2-38: Illustration of judgement distributions of stimulus and category boundaries.

Torgerson's law of categorical judgement is defined by Equation 2-49. However, it cannot be solved in its complete form, since the number of unknown variables exceeds the number known. To solve this problem, some assumptions must be made. In practice, the dispersions σ_i and σ_j , and the correlation ρ_{ij} are seldom known. Therefore, σ_i and σ_j , are assumed to be the same and ρ_{ij} is taken to have a value of zero. This is known as Thurstone's Case V for the law of comparative judgements (Thurstone, 1959). This allows the scale values to be determined using Equation 2-50.

$$T_j - S_i = z_{ij} \sigma \sqrt{2} \quad \text{Equation 2-50}$$

2.9.2 Magnitude Estimation

Magnitude estimation was first used by Richardson and Ross (1930) and then elaborated by Stevens (1953). This method is similar to the category judgement scaling method in many ways. The most notable differences are first, the number of categories is much larger (*e.g.*, any positive number) and secondly, this method relies on responses of observers so as not to invoke assumptions about initial categories which possibly cover all observer responses against presented stimuli. In the magnitude estimation method, observers are asked to make direct numerical estimations of the perceived magnitude produced by various stimuli regarding the attribute under interest. It has to be mentioned that although it is called magnitude estimation, it is not the magnitudes that are specified. What is specified is a ratio; the ratio of one perception with respect to another implicit or explicit perception (Bartleson, 1984). There are two main ways of applying the magnitude estimation technique. One way is that observers are presented with a reference stimulus and told that the perception it produces has a certain defined numerical value, such as 10. Subsequently, other stimuli are presented and observers are asked to make judgements by comparison with the reference stimulus (the ratio between the test stimulus and a reference stimulus). Therefore, the estimations are constrained by the reference stimulus. In the other version, an unconstrained method is used by omitting the reference stimulus. The various stimuli are randomly presented to observers who assign numbers in proportion to the perceived magnitude of the perception.

In order to determine scale values for the stimuli, the raw data collected from observers are usually averaged by taking their geometric mean \bar{O} using Equation 2-51 or sometimes by taking the median for each stimulus.

$$\bar{O}_i = (\prod O_i)^{1/n} \quad \text{Equation 2-51}$$

where O is the observer's response, i stands for a particular stimuli and n represents the number of the observations.

The rationale for using the geometrical mean is that the dispersions tend to be normally distributed over the logarithms of responses and also standard deviations tend to increase linearly with the mean. This means that equal intervals on logarithmic scales correspond to equal ratios. The arithmetic mean is rarely used because it may be excessively influenced by a few unrepresentative high judgements. One problem of taking the geometric mean is caused by zero responses because a zero response from

just one observer makes the geometric mean have a value of zero. A fundamental property of scale is that it has an absolute zero. In practice, this was not expected to be encountered, but there are many exceptions. An alternative method of computing the geometric mean of the data with zero responses is to add a small constant α to each observer's response. Then the constant is subtracted from the computed geometric mean as shown in Equation 2-52 (Engeldrum, 2000). Appropriate values of this constant should be very small, say 0.001, otherwise it can seriously affect the mean especially if there are many zero responses.

$$\bar{O}_i = (\prod O_i + \alpha)^{1/n} - \alpha \quad \text{Equation 2-52}$$

After the collection of many experimental results using the ratio method, to various stimuli such as brightness, loudness smell and so on, Stevens (Stevens, 1961) concluded that the central tendency of the numerical response of observers O is approximately a power function of the stimulus intensity I with exponent γ and constants α and β (which may be applied depending on several factors) as given in Equation 2-53. This is widely known as *Stevens' power law* (the original law is without the constant α).

$$O = \alpha I^\gamma + \beta \quad \text{Equation 2-53}$$

Thus the results of magnitude estimation experiments are analysed.

2.10 Statistical Measures

Various experimental data obtained in this study were analysed using a number of statistical techniques. Descriptive statistics were used to describe the basic features of the data such as the mean, median and range of the data. The *correlation coefficient* (R) and the *coefficient of determination* (R^2) were frequently used to indicate the degree of a linear relationship between two data sets. In addition to these qualitative measurements, the *root mean square error* (*RMSE*) which is categorised as a quantitative method, and the *coefficient of variation* (*CV*) were also utilised to assess a relationship between data sets. In some cases, to describe the probabilistic properties of data, a *95 % confidence interval* was calculated and the *Wilcoxon matched-pairs signed-rank test* performed (Kothari, 2005; Upton & Cook, 1996). The methods utilised in this study are introduced in the following sections.

2.10.1 Correlation Coefficient & Coefficient of Determination

The correlation coefficient (R , also called the Pearson product-moment correlation coefficient) measures the strength and direction of a linear relationship between two data sets. The range of the correlation coefficient is from -1.0 to 1.0 , where -1.0 represents a perfect negative correlation and 1.0 represents a perfect positive correlation. The value of zero indicates a non-linear relationship between the two data sets. The formula for computing the correlation coefficient is given in Equation 2-54.

$$R = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{[\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2]^{0.5}} \quad \text{Equation 2-54}$$

where x_i and y_i are individual values of two data sets for stimulus i , and \bar{x} and \bar{y} are the mean values of the two data sets.

The coefficient of determination (R^2) represents the proportion of variability between data sets. Although the value of R^2 can be obtained practically from the square of R , the original idea of R^2 was to determine the total variance between two data sets, rather than the linear correlation between two sets. R^2 is often used in regression analysis as a measure of goodness-of-fit which indicates how well data fits to an estimated model. R^2 values range from 0 to 1; a larger value indicates better correlation; and zero represents no correlation.

R and R^2 conveniently describe the correlation of two data sets with a single number. Unfortunately, high values of R and R^2 do not always guarantee high correlation. Examples are given using Figure 2-39. All data sets plotted in each graph in Figure 2-39 have similar R and R^2 values. However, it can be seen that they exhibit very different features in their distribution. R and R^2 values are appropriate measures to describe the data sets plotted in Figure 2-39 (a). However, the data sets given in Figure 2-39 (b) might be criticised as being a poor correlation. The plotted graph in Figure 2-39 (c) suggests that a non-linear correlation may better describe the characteristics of the data sets. Thus, these examples suggest that the data analysis should not rely on those numerical methods alone and, as a minimum, the data should be plotted, or another measure found.

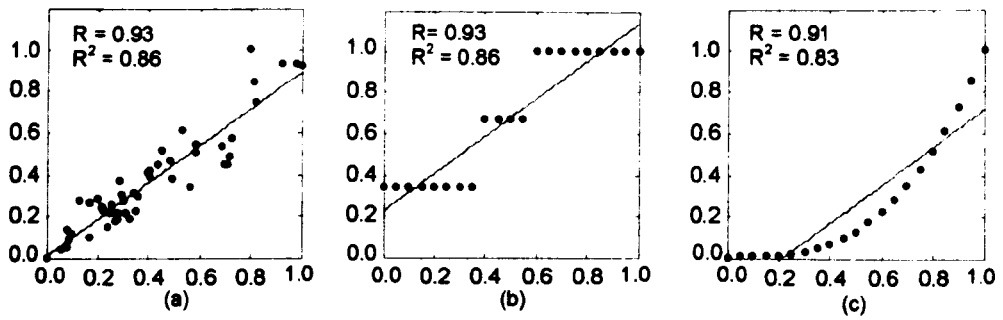


Figure 2-39: Examples of the relationships between data sets and the correlation coefficient or the coefficient of determination.

2.10.2 Root Mean Square Error (RMSE)

Root mean square error (*RMSE*) measures an average magnitude of errors. The range of the *RMSE* is from zero to infinity. A minimum *RMSE* value of zero indicates that two data sets perfectly agree with each other. A maximum *RMSE* value, which is an indication of disagreement, depends on the scale range of both data sets. The *RMSE* is determined by Equation 2-55.

$$RMSE = \sqrt{\frac{\sum_i (x_i - y_i)^2}{N}} \quad \text{Equation 2-55}$$

where x_i and y_i are individual values of two data sets on stimulus i and N is the number of stimuli.

2.10.3 Coefficient of Variation (CV)

The coefficient of variation is another measure of the variability between two data sets. If they agree perfectly, all the data points lie on a 45° line when they are plotted. The wider the scatter of the data points, the poorer the agreement. A value of zero indicates a perfect correlation between two data sets. If two data sets have different units, it is necessary to first transform the units using a scaling factor SF . CV can be expressed by Equation 2-56.

$$CV = \frac{100}{y} \sqrt{\frac{\sum_i (y_i - SFx_i)^2}{N}} \quad \text{Equation 2-56}$$

$$SF = \frac{\sum_i (x_i y_i)}{\sum_i x_i^2}$$

where N is the total number of stimulus, x_i and y_i are individual values of two data sets on stimulus i , and \bar{y} is a mean value of y_i .

2.10.4 Confidence Intervals

The confidence interval of a population mean gives an indication of the uncertainty in the estimate of the true mean. A small interval indicates that the estimation is more precise. The choice of confidence coefficient is somewhat arbitrary: a 90 %, 95 % or 99 % confidence interval is often used with 95 % being the most common. This can be computed using Equation 2-57.

$$95\% \text{ Confidence Interval} = \overline{SV} \pm t_{95\%(N-1)} \frac{\sigma}{\sqrt{N}} \quad \text{Equation 2-57}$$

where \overline{SV} is the mean of the data, N is the number of observations, such that $N-1$ indicates the number of degrees of freedom, σ is the standard deviation and $t_{95\%(N-1)}$ represents the upper critical value of the t -distribution. When the sample size is sufficiently large, or even though sample size is small when the distribution of the population can be assumed to be normal, 1.96 is used instead of $t_{95\%(N-1)}$ (95 % of the area of a normal distribution is within 1.96 standard deviations of the mean). In this study, the sample sizes (the number of the observations per sample) are not large enough and the distributions cannot be assumed to be normal. Therefore, t -distribution is applied.

2.10.5 The Wilcoxon Matched-Pairs Signed-Rank Test

Statistical analysis is often used to infer the probability that any differences between data are caused by chance. The Wilcoxon matched-pairs signed-rank test (or the *Wilcoxon signed-rank test*) is a nonparametric analogue to the *paired t-test* which is most commonly used. Unlike the paired t -test, the Wilcoxon matched-pairs signed-rank test does not assume that the data distribution can be approximated by the normal distribution. This test takes account of the magnitude of the difference between data sets based on their median values. Therefore, this method tests the null hypothesis (H_0) against one of the alternative hypotheses (H_1 , H_2 or H_3):

H₀: the median of two data sets are same

H₁: the median of two data sets are different

H₂: the median of one data set is larger than the other data set

H₃: the median of one data set is smaller than the other data set

A worked example is given in Table 2-1. When this test is applied, the differences (denoted in Table 2-1 as d_i) between each pair of values (X_i, Y_i) can first be determined. If the differences are zero, they should be discarded from the calculations. The rank is then assigned to each difference without regard to sign. Consequently, the original signs of each difference are restored to corresponding ranks (*Rank*). When two or more pairs have the same difference values, the assigned ranks to such pairs are averaged and given the same rank. For example, if two pairs have rank values of 3, the rank of 3.5 is assigned for each rank value (*i.e.*, $(3 + 4)/2 = 3.5$) and then the next largest difference should be 5. Then, W^+ and W^- are computed as the sums of the positive and negative ranks respectively. The smaller of two W^+ and W^- in absolute values can be assigned as the test statistics T . When the number of pairs n , after discarding the pairs which have zero difference, is equal to or less than 25 ($n \leq 25$), the critical value of T can be found according to the Wilcoxon signed-rank statistic distribution in order to evaluate the hypothesis of whether to accept or reject the null hypothesis at the desired level of significance (5 % error rate is commonly used). If the number of pairs exceeds 25 ($n > 25$), the test statistic T is evaluated using the normal distribution with a mean of μ and a standard deviation σ , as given in Equation 2-58 and Equation 2-59 respectively.

$$\mu = \frac{n(n+1)}{4} \quad \text{Equation 2-58}$$

$$\sigma = \sqrt{\frac{n(n+1)(2n+1)}{24}} \quad \text{Equation 2-59}$$

The test statistic z can be described by Equation 2-60.

$$z = \frac{n(n+1)/4 - T}{\sqrt{n(n+1)(2n+1)/24}} \quad \text{Equation 2-60}$$

The appropriate critical value is defined according to the normal distribution.

Consider the example shown in Table 2-1 for the hypothesis,

H_0 : the median of two data sets are the same

H_1 : the median of two data sets are different

Since the test statistic $T = 11$ is greater than the critical two-tailed 0.05 value of $T_{0.05} = 8$, the null hypothesis is accepted. This indicates that the probability value according to the test statistic is more than 5 % ($p > 0.05$).

Table 2-1: A work example for the Wilcoxon matched-pairs signed-rank test.

x_i	3	4	6	8	5	14	7	2	13	19
y_i	4	6	3	5	1	10	2	7	8	9
d_i	-1	-2	3	3	4	4	5	-5	5	10
Rank	-1	-2	3.5	3.5	5.5	5.5	8	-8	8	10
$W+ = 3.5+3.5+5.5+5.5+8+8+10 = 44$										
$W- = 1+2+8 = 11$										

2.11 Summary

The present study deals with various issues for studying the human visual system, colorimetry and appearance. The available literature is large. Hence, only that directly related to the intended study has been reviewed. The review shows that the appearance of metallic coatings consisting of qualities such as colour and gloss belonging to macro appearance, and many visual texture attributes belonging to micro appearance. Colour is obviously important when considering the appearance, and a relatively high number of study have been carried out. However, there seems to be less information about visual texture. Because of the physical complexity of gonioapparent metallic coatings and the subjective nature of texture attributes, even the definitions associated with visual texture remain ill-defined. Moreover, although many methods to quantify texture have been introduced, there is no unique method which can be applied for all textured surfaces. These uncertainties give motivation for the investigations into the perception of coarseness and glint. Also, with the great expectation of digital image acquisition and visual assessments of appearance attributes on computer displays, the current project is considered to be very challenging.

Chapter 3

Assessing and Modelling Coarseness using Physical Samples

3.1 Introduction

This chapter describes a study aimed to characterise perceptual coarseness of metallic-coating panels. The coarseness of metallic-coating panels was perceptually scaled and a computational model capable for predicting the perceptual coarseness was developed.

“Coarseness” is a general term that can be used to describe the human response to perceived texture. In the literature, it has been referred as a key feature in the analysis of texture and also as an important appearance attribute of metallic coatings (see Section 2.5.1.2). In terms of human recognition, it is not difficult to identify, but it is not trivial to quantify since the appearance of coarseness is generally size and resolution dependent (see Section 1.2). In addition, illumination and viewing geometry is crucial to texture appearance; especially for gonioapparent materials such as the metallic coatings used in this study (see Section 2.5). In order to characterise the perceptual coarseness of metallic coatings, it is necessary to establish an identifying condition where the coarseness could be observed. Therefore, the following sections describe a preliminary investigation to find appropriate illumination and viewing geometry for observation of the coarseness of the metallic coating and the visual assessments for scaling the perceptual coarseness of a series of metallic-coating panels under defined conditions as well as subsequent evaluations of the results at first. Then, a computational model which was developed for predicting the perceptual coarseness of the metallic-coating panels is introduced. Finally, the performance of the developed model which is evaluated by comparing the model prediction with the perceptual coarseness as illustrated in Figure 3-1, is described. Before starts the characterisation of coarseness, instruments used for colour measurement in this study and their performance are introduced.

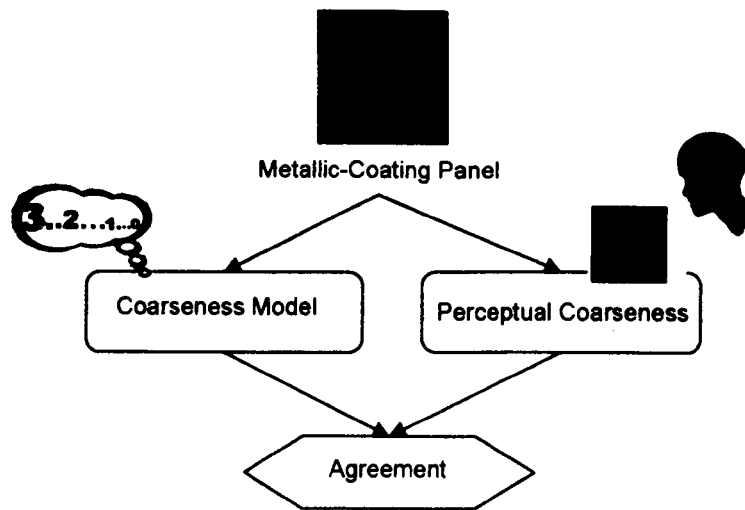


Figure 3-1: A flowchart for the study of the present chapter.

3.2 Colour Measurement Instruments

Colour measurement is essential in the present study to specify colour of metallic-coating panels used and also to describe experimental conditions such as illumination conditions. Two colour measuring instruments were employed. A *Minolta CS1000 tele-spectroradiometer* (TSR) was used for measuring spectral power distribution (SPD). A GretagMacbeth ColorEye 7000A spectrophotometer (CE7000A) was utilised for surface reflectance measurement. Variability in the measurements of colour can be caused by fluctuation in the measurements made using a measuring instrument, in a target to be measured and in the environmental conditions (e.g., illumination). Hence, it is important to quantify the variability of the instruments. In the following sections, the specification of the instruments and their performance are described.

3.2.1 Minolta CS1000 Tele-Spectroradiometer

A Minolta CS1000 tele-spectroradiometer (TSR) is configured with an SLR (single lens reflex) optical system for targeting an object and a high-resolution photodiode array for capturing the signal. The measurement is made over the visible spectrum from 380 to 780 nm with a fixed interval of 5 nm. The minimum measuring area is 7.9 mm in diameter.

The performance of the TSR was investigated in terms of repeatability. A tungsten light source in an integrating sphere (Bentham SRS8Q Spectral

Radiance/Luminance Standard) was measured five times in 30 minutes. The light source and the TSR were allowed 30 minutes warm-up time before the measurements. The measurement results are given in Table 3-1 in terms of CIE XYZ tristimulus values and their standard deviations. It should be aware that the colorimetric data which were obtained using this instrument in this study, were including errors such as repeatability and accuracy errors. This data set was provided by Colour and Imaging Group at the Department of Colour Science in the University of Leeds in 2007.

Table 3-1: Repeatability test of Minolta CS1000 tele-spectroradiometer.

Measurement	X	Y	Z
1 st	97.81	86.07	24.89
2 nd	97.83	86.09	24.81
3 rd	97.81	86.06	24.87
4 th	97.81	86.06	24.87
5 th	97.80	86.05	24.81
Standard Deviation	0.01	0.02	0.04

3.2.2 GretagMacbeth ColorEye 7000A Spectrophotometer

A GretagMacbeth ColorEye 7000A spectrophotometer (CE7000A) has measurement geometry of d/8 provided by an integrating sphere (see Section 2.4.2). It measures spectral reflectance from 360 to 750 nm at 10 nm interval. A pulsed xenon light source with high intensity ensures accurate readings, even for dark or highly saturated colours. Measurement can be made with specular component included or excluded by adjusting a motorised spectral insert, and also a motorised UV control is for adjusting the UV content. An aperture size for measurement can be selected from four different sizes: large (25.4 mm circular), medium (15 mm circular), small (8 mm circular) and very small (3 × 8 mm). In this study, the measurements were made with the setting of the specular component excluded using a small aperture size.

Accuracy of the CE7000A was investigated based on a set of BCRA-NPL glossy ceramic tiles provided by the NPL. Ten tiles were consecutively measured two times and the mean values were compared with the calibration data provided for each tile. The results are reported in terms of CIELAB ΔE^*_{ab} in Table 3-2. The CIELAB coordinates were computed from the reflectance assuming the D65 illuminant and using the CIE 1964 standard colorimetric observer. Repeatability was tested by measuring the four of the ceramic tiles successively ten times at the same position. The results are

summarised in Table 3-3. The mean and maximum errors were ΔE^*_{ab} values of 0.42 and 1.02 respectively. The repeatability was found in a range of mean ΔE^*_{ab} values from 0.004 to 0.047. These accuracy and repeatability errors were contained in the measurements carried out in this study with this instrument.

*Table 3-2: Accuracy test of the CE 7000A in terms of ΔE^*_{ab} of 10 tiles between the measurement data using the CE7000A and the calibration data provided by the NPL.*

Tile	ΔE^*_{ab}
Pale Grey	0.30
Mid Grey	0.21
Diffuse Grey	0.16
Deep Grey	0.26
Deep Pink	0.25
Red	0.75
Orange	0.75
Green	0.32
Diffuse Green	0.17
Green	1.02
Mean	0.42
Maximum	1.02
Minimum	0.17

*Table 3-3: Repeatability test of the CE 7000A in terms of mean, maximum and minimum ΔE^*_{ab} values of 10 successive measurements.*

Tile	Mean ΔE^*_{ab}	Maximum ΔE^*_{ab}	Minimum ΔE^*_{ab}
Mid Grey	0.004	0.007	0.002
Red	0.047	0.079	0.010
Green	0.016	0.031	0.003
Deep Blue	0.026	0.050	0.006

3.3 Visual Assessments of Coarseness on Metallic-Coating Panels

The following sections introduce a method to characterise coarseness perceptually. Appropriate illumination and viewing geometry to observe coarseness and a method to quantify coarseness of metallic-coating panels are discussed as well as the results thus obtained.

3.3.1 Samples

A set of 156 metallic-coating panels produced by Akzo Nobel were used as samples. These panels were made by mixing solid-colour pigments and aluminium flakes in different proportions. There were 6 grey colour panels, 50 blue colour panels, 50 green colour panels and 50 purple colour panels. In this experiment, a part of the panel, the size of 8 × 8 cm, was presented to observers. Images of one sample from each colour group are given in Appendix I as examples. Note that the appearance of the images may not be the same as that perceived for the physical samples.

3.3.2 Viewing Conditions

The overall appearance of any object consists of a combination of various attributes (see Section 1.2). One attribute might appear stronger than the others under one viewing condition but another attribute might be more predominant in a different viewing condition. The coarseness of metallic coatings is clearly visible under diffuse illumination (see Section 2.5.1.2). Although the coarseness is seen under different conditions, for example, directional illumination, observers can be distracted by undesired attributes such as specular reflection or gloss. ASTM (2006) defines diffuse illumination as “an extended-area source”. According to this definition, a typical viewing cabinet as illustrated in Figure 3-2, which is often used to make visual assessments, provides such extended-area illumination. However, it was found that this type of viewing cabinet was not appropriate for the coarseness assessments of the metallic-coating samples, as a result of a visual inspection. Since the cabinet has tubular lamps in the ceiling, effects which are similar to a directional light were observed at some viewing angles. Therefore, in order to carry out the visual assessments under stable diffuse illumination condition, a DigiEye® viewing cabinet (Luo and Cui *et al.*, 2001) which is especially made to provide diffuse illumination, was used for this experiment.



Figure 3-2: A typical viewing cabinet used for visual assessment.

3.3.2.1 Experimental Settings

The experimental settings are illustrated in Figure 3-3 (a) and (b). A DigiEye® viewing cabinet was used for this experiment, incorporating diffuse illumination from two light sources (CIE illuminant D65 simulators) covered by diffusing filters and a flat base to present samples, as shown in Figure 3-3 (a). Hence, the illumination and viewing geometry was $d/0$. Two samples of metallic-coating panels which were used as a reference and a test sample, were placed on a uniform mid-grey background on the bottom of a cabinet as illustrated in Figure 3-3 (b). The visual assessment was conducted in a darkened room. Each observer was seated on a chair in front of the cabinet placed on the floor. An observer looked down onto the samples from the viewing window on the top of the cabinet. The distance from the observer's eye to the sample was approximately 54 cm which was considered an appropriate distance to assess micro appearance (see Section 2.5.1). It was slightly farther than 25 cm which is suggested by McCamy (1998) for the visual texture observations, but similar to the distance of 50 cm used in the coarseness assessment of coatings by Akzo Nobel (Kirchner and Kieboom *et al.*, 2007). The two samples were separated by 10 cm from each other across the centre area in order to avoid observers seeing a reflected mirror image which can be seen on the sample if it is placed directly under the viewing window, since the metallic coatings have a high gloss surface.

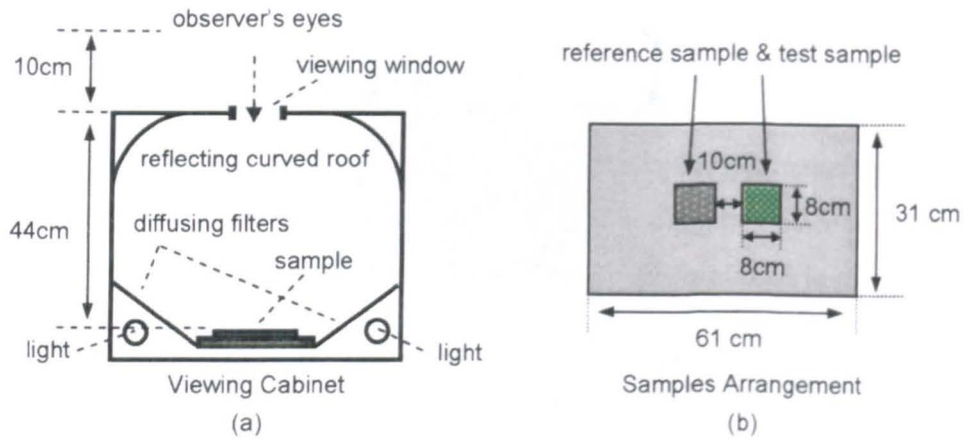


Figure 3-3: (a) A schematic diagram of a DigiEye® Viewing Cabinet. The samples are placed on the base. Two light sources (D65 simulators) are positioned at two bottom corners at each side and both emit light to the walls. The two top corners have curved surfaces to reflect light uniformly onto the samples. An observer looked down onto the sample from the viewing window. (b) The arrangement of the samples as presented to the observers.

The SPD of the light source was specified by measuring the SPD of the light reflected from a white ceramic tile using the TSR mounted on the top of the cabinet. The SPD is given in Figure 3-4. The luminance level of the light reflected from the tile was 167.8 cd/m^2 . The CIE XYZ tristimulus values and the CIE (x, y) chromaticity coordinates in terms of the CIE 1964 standard colorimetric observer were (X, Y, Z) equal to (171.0, 186.7, 217.1) and (x, y) equal to (0.2974, 0.3248) respectively. The chromaticity coordinates can be compared to the CIE 1964 (x, y) chromaticity coordinates of 0.3138 and 0.3310 for the CIE Illuminant D65 (CIE, 2004b). The uniform mid-grey background of the samples had CIELAB (L^* , a^* , b^*) values equal to (56.01, 0.3431, 1.3727) according to the measurement using the TSR at the centre of the viewing field.

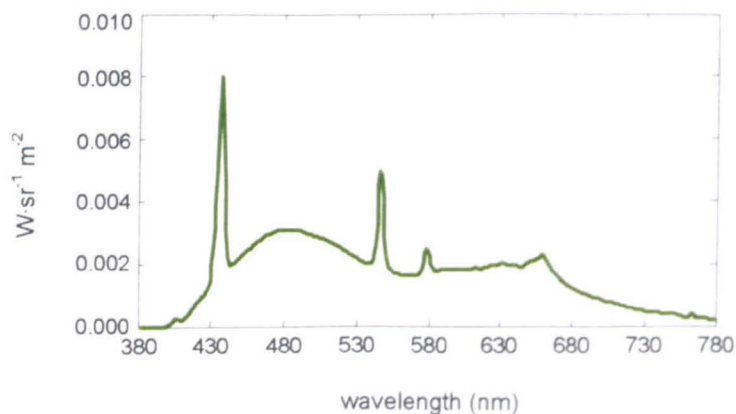
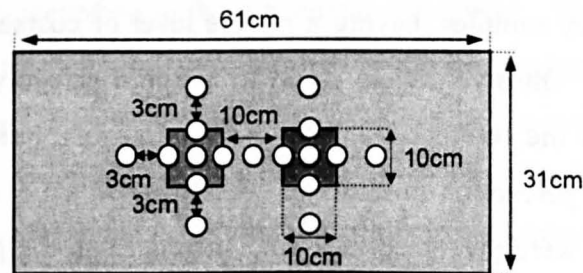


Figure 3-4: SPD of the light source specified by measuring SPD of a white ceramic standard tile using a Minolta CS1000 tele-spectroradiometer (TSR).

Spatial uniformity of the illumination was investigated by measuring the grey background using the TSR at 17 points shown in Figure 3-5 (a). It was necessary for these measurements to tilt the TSR except at the centre. The TSR could not be placed directly above each point, because of the structure of the cabinet which only had a small viewing window in the top. Figure 3-5 (b) shows the luminance in cd/m^2 at each point corresponding to Figure 3-5 (a). The values of CIELAB ΔE^*_{ab} colour difference (see Section 2.3.6) between the centre and each point are given in Figure 3-5 (c). It can be seen that the luminance levels at the areas where the samples were presented were adequately uniform: The maximum and minimum luminance at the sample areas were 38.41 cd/m^2 and 37.91 cd/m^2 respectively. The difference between these values was only 0.5 cd/m^2 . The maximum colour difference at the sample areas was equal to 0.83 and the minimum to 0.23. These errors can be considered to be insignificant.



(a)

		38.49				38.78		
		38.07				38.07		
38.17	38.15	37.94	37.91	38.25	38.41	37.93	37.97	37.97
		38.18				38.13		
		38.30				38.78		

(b)

		0.48				0.63		
		0.40				0.52		
0.62	0.32	0.30	0.58	0	0.23	0.67	0.63	0.60
		0.83				0.36		
		0.57				0.62		

(c)

Figure 3-5: Spatial uniformity evaluation of the illumination. (a) Locations of measurements in uniformity evaluation. (b) Luminance in cd/m^2 at each measurement point. (c) CIELAB ΔE^*_{ab} colour differences between the centre and each point.

3.3.3 Scaling Perceptual Coarseness

Ten observers, including four females and six males aged between 24 and 34, participated in the visual assessments for scaling perceptual coarseness of the metallic-coating panels. All observers were either students or members of staff at the University of Leeds. They all passed the Ishihara visual test as observers with normal colour vision.

Each observer commenced an observation session by adapting to the viewing field which was the mid-grey background in the cabinet. Unlike length or weight, there is not any unit for scaling perceptual coarseness. Therefore, even though the concept of the coarseness is relatively easy to understand, observers may find it difficult to scale coarseness without any reference. This problem was solved by applying a Categorical judgment scaling method (see Section 2.9.1) with a reference sample so as to calibrate observers on the same scale. Two samples were presented for each assessment in the cabinet. One was the reference sample and the other a test sample (see Figure 3-3). One of the six grey samples, having a middle level of coarseness, was chosen as the reference sample. Observers were asked to assign a category for the test sample by comparing it with the reference sample, whose category was set at a value of five, according to their perception in terms of coarseness on a 1-9 scale as shown in Table 3-4. All samples were presented in a random order and the position of the reference sample was also randomly selected on either the right- or left-hand side of the test sample during each assessment so as to prevent any systematic and experimenter bias (Dean & Voss, 1999). A training session was conducted before the main session. In the training session, each observer was asked to assess samples including the five grey colour samples and ten random samples from the other coloured samples. The training session was to help the observers get used to the experimental procedure and therefore the results were not included in the overall experimental analysis and the observers were also informed that it was a training session. After the training session, the main session involving the assessments of all the 155 samples was conducted. To test repeatability, each observer carried out the main session twice. A total of 20 main sessions which gave a total of 3100 categorical judgements (10 observers × 2 sessions × 155 samples) were needed.

Table 3-4: 1-9 categories used for the visual assessment.

Category	Category Description
Category 1	Extremely Fine
Category 2	Very much Fine
Category 3	Moderately Fine
Category 4	Slightly Fine
Category 5	Reference Sample
Category 6	Slightly Coarse
Category 7	Moderately Coarse
Category 8	Very much Coarse
Category 9	Extremely Coarse

3.3.4 Data Analysis and Results

3.3.4.1 Scale Value

The raw experimental data consisted of the category numbers assigned by the observers to each sample according to the perceptual coarseness. The simplest way to analysis these data is to take an arithmetic mean over the results from all observers to obtain a mean scale value for each sample. This is known as the *mean-category value method* (see Section 2.9.1) and assumes that the observers are capable of keeping the intervals between category boundaries psychologically equal. However, there is always doubt about an observer's ability to categorise samples into an equal-interval of the categories. Thus, a more sophisticated technique known as the *categorical-judgement method* is preferred (see Section 2.9.1). This transforms scaled data to an equal-interval scale. A worked example is given in Table 3-5 (Bartleson, 1984; Luo, 2002).

1. An $m \times n$ **frequency matrix** was constructed regarding m samples (stimuli) and n categories in a categorical judgement assessment. Each entry shows the frequency for a sample being judged as being in a specific category.
2. An $m \times n$ **cumulative frequency matrix** was constructed in which each entry shows the frequency of a sample judged to be below a given category.
3. An $m \times n$ **cumulative probability matrix** was obtained by divided each entry in the cumulative frequency matrix by the number of observations.
4. An $m \times n$ **LG matrix** (logistic function values) is obtained from the cumulative frequency matrix using Equation 3-1. This function can be used to estimate z-scores in Step 5.

$$LG = \ln\left(\frac{CF + \alpha}{N - CF + \alpha}\right) \quad \text{Equation 3-1}$$

where CF represents the cumulative frequency matrix, N is the number of the observations and α is an arbitrary additive constant (0.5 was suggested by Bartleson (1984) and was used in this study).

5. Z-scores can be obtained from the cumulative probability matrix as probability is the area (proportion) under the normal distribution curve. According to a property of the normal distribution, for those having probability values of 0 or 1, the z-scores are $-\infty$ or ∞ respectively. Therefore, in both cases, an $m \times n$ z-score matrix was estimated from the LG values using a scaling coefficient α (and if necessary a coefficient β) as given in Equation 3-2, which was calculated using linear regression between the valid z-scores and corresponding LG value.

$$z \text{ score} = \alpha LG + \beta \quad \text{Equation 3-2}$$

6. An $m \times (n - 1)$ **difference matrix** between adjacent columns was calculated followed by calculation of the mean for each column.
7. **Category boundaries** were determined by setting the origin (in the worked example, it is the boundary between Category 1 and 2) to zero and adding adjacent mean values from the different matrix.
8. An $m \times (n - 1)$ **scale value matrix** was calculated by subtracting each entry of the z-score matrix from the corresponding category boundaries. The mean values of each row represent the coarseness of the samples as their scale value.

Table 3-5: A worked example of the categorical-judgement method. In this example, the number of the sample, $m = 3$, the category, $n = 7$ and the total observations, $N = 8$, by 4 observers (twice by each observer).

Category	1	2	3	4	5	6	7
Observer 1							
Sample 1	0	0	0	2	0	0	0
Sample 2	0	0	0	1	1	0	0
Sample 3	1	1	0	0	0	0	0
Observer 2							
Sample 1	0	0	0	0	2	0	0
Sample 2	0	0	0	1	1	0	0
Sample 3	0	0	2	0	0	0	0
Observer 3							
Sample 1	0	0	0	0	2	0	0
Sample 2	0	1	1	0	0	0	0
Sample 3	0	2	0	0	0	0	0
Observer 4							
Sample 1	0	0	0	0	1	1	0
Sample 2	0	0	0	1	1	0	0
Sample 3	0	0	1	1	0	0	0

Category	1	2	3	4	5	6	7
Frequency matrix							
Sample 1	0	0	0	2	5	1	0
Sample 2	0	1	1	3	3	0	0
Sample 3	1	3	3	1	0	0	0
Cumulative frequency							
Sample 1	0	0	0	2	7	8	8
Sample 2	0	1	2	5	8	8	8
Sample 3	1	4	7	8	8	8	8
Cumulative probability matrix							
Sample 1	0	0	0	0.25	0.875	1	1
Sample 2	0	0.125	0.25	0.625	1	1	1
Sample 3	0.125	0.5	0.875	1	1	1	1
LG matrix							
Sample 1	-2.833	-2.833	-2.833	-0.956	1.609	2.833	2.833
Sample 2	-2.833	-1.609	-0.956	0.452	2.833	2.833	2.833
Sample 3	-1.609	0.000	1.609	2.833	2.833	2.833	2.833
z-score matrix							
Sample 1	-2.020	-2.020	-2.020	-0.680	1.149	2.023	2.023
Sample 2	-2.020	-1.147	-0.680	0.324	2.023	2.023	2.023
Sample 3	-1.147	0.001	1.149	2.023	2.023	2.023	2.023
Difference matrix							
Sample 1	0.000	0.000	1.340	1.830	0.873	0.000	
Sample 2	0.873	0.467	1.004	1.699	0.000	0.000	
Sample 3	1.148	1.148	0.873	0.000	0.000	0.000	
Mean	0.674	0.538	1.072	1.176	0.291	0	
Boundary	0.000	0.674	1.212	2.284	3.460	3.751	3.751
Scale value matrix							
							Scale Value (Mean)
Sample 1	2.020	2.694	3.232	2.965	2.311	1.729	2.492
Sample 2	2.020	1.821	1.892	1.960	1.438	1.729	1.810
Sample 3	1.147	0.672	0.062	0.262	1.438	1.729	0.885

The raw experimental data were analysed using the two methods: the mean-category value method and the categorical-judgement method. These methods were compared by plotting two sets of results as shown in Figure 3-6 where the mean-category value method is plotted on the horizontal axis and the categorical-judgement method on the vertical axis. As can be seen in Figure 3-6, the coefficient of the determination value (R^2) of 0.9966 indicates that the results from both methods are well correlated. Frequently, it is found that a simple mean-category value method yields scale values that are very similar to those determined by a categorical-judgement method (Bartleson, 1984; Han, 2006). This high level of correlation indicates that the

observers could follow the instructions with the respect to the equal-interval properties of the category scales with a high degree of precision. Consequently, the scale values derived from the simple mean-category value method were used to represent the perceptual coarseness of the samples scaled by the observers. The scale values of all the samples are given in Appendix II.

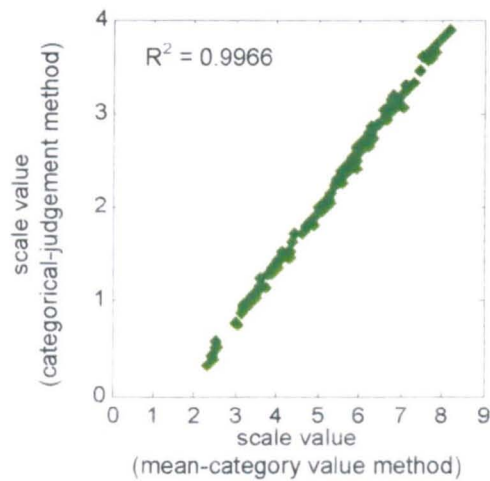


Figure 3-6: Comparison of the scale values derived from the mean-category value method and the categorical-judgement method for all samples.

3.3.4.2 Observer Variability

In this section, the uncertainty in the experimental results is determined by observer variability. Observer variability was investigated from two aspects: *observer repeatability* and *observer accuracy*, sometimes called *intra-observer agreement* and *inter-observer agreement* respectively. Observer repeatability indicates how well the experimental result agrees with a result replicated by the same observer. Observer accuracy investigates how well individual observers agree with the mean experimental result. Thus, for the repeatability investigation, the raw data in the first session of the coarseness scaling for the 155 samples was compared with that in the second session for individual observers. The raw data means the category data assigned by each observer for each sample. Observer accuracy was investigated by comparing the raw data of each session with the mean-category values. These comparisons were carried out in terms of the *coefficient of determination* (R^2) and the *coefficient of variation* (CV) (R^2 and CV; see Sections 2.10.1 and 2.10.3) and the results are summarised in Table 3-6 and Table 3-7. In addition, a 95 % confidence interval was computed to obtain a measure of the performance of the observers (see Section 2.10.4).

Table 3-6: Observer repeatability and the associated mean, median, maximum and minimum values of all the samples and the individual grey, blue, green and purple samples in terms of R² and CV.

Observer	R ²					CV				
	All	Grey	Blue	Green	Purple	All	Grey	Blue	Green	Purple
1	0.70	0.74	0.36	0.40	0.35	16	22	14	13	25
2	0.79	0.92	0.36	0.32	0.57	17	20	15	12	29
3	0.55	0.81	0.19	0.36	0.36	18	21	19	14	20
4	0.66	0.92	0.27	0.47	0.19	21	13	19	11	39
5	0.53	0.90	0.17	0.30	0.17	19	13	17	15	27
6	0.74	0.96	0.41	0.26	0.50	16	8	19	15	14
7	0.77	0.95	0.42	0.42	0.46	12	8	10	8	19
8	0.67	0.95	0.20	0.30	0.21	24	12	20	13	54
9	0.75	0.98	0.43	0.57	0.57	19	9	20	15	25
10	0.74	0.89	0.38	0.28	0.14	17	12	14	12	32
Mean	0.69	0.90	0.32	0.37	0.35	18	14	17	13	28
Median	0.72	0.92	0.36	0.34	0.36	17	12	18	13	26
Max	0.79	0.98	0.43	0.57	0.57	24	22	20	15	54
Min	0.53	0.74	0.17	0.26	0.14	12	8	10	8	14

Table 3-7: Observer accuracy for each session and the associated mean, median, maximum and minimum values of all the samples and the individual grey, blue, green and purple samples in terms of R² and CV.

Observer & Session		R ²					CV				
		All	Grey	Blue	Green	Purple	All	Grey	Blue	Green	Purple
1	1	0.82	0.91	0.64	0.66	0.58	11	14	9	10	18
	2	0.88	0.98	0.50	0.72	0.76	18	20	16	14	26
2	1	0.69	0.96	0.60	0.58	0.56	16	8	17	13	20
	2	0.83	0.98	0.59	0.68	0.65	12	10	11	8	19
3	1	0.68	0.75	0.45	0.45	0.45	16	19	13	11	26
	2	0.84	0.93	0.56	0.53	0.69	16	11	19	13	17
4	1	0.83	0.98	0.61	0.54	0.59	13	10	8	8	24
	2	0.81	0.97	0.49	0.32	0.51	18	8	15	12	33
5	1	0.82	0.93	0.61	0.76	0.73	15	14	15	12	21
	2	0.81	0.95	0.54	0.61	0.23	13	8	10	7	27
6	1	0.88	0.93	0.51	0.72	0.72	11	19	11	7	15
	2	0.88	0.97	0.66	0.55	0.70	15	8	13	13	21
7	1	0.78	0.88	0.47	0.65	0.71	13	17	11	9	20
	2	0.84	0.91	0.44	0.74	0.44	17	15	16	8	31
8	1	0.74	0.58	0.31	0.66	0.42	14	22	13	12	18
	2	0.86	0.99	0.66	0.56	0.70	10	5	9	8	14
9	1	0.88	0.98	0.60	0.65	0.63	11	8	10	8	19
	2	0.82	0.99	0.45	0.58	0.55	11	5	10	9	18
10	1	0.86	0.99	0.76	0.81	0.78	15	13	16	12	20
	2	0.81	0.96	0.58	0.48	0.27	14	7	13	10	22
Mean		0.82	0.93	0.55	0.61	0.58	14	12	13	10	21
Median		0.83	0.96	0.57	0.63	0.61	14	10	13	10	20
Max		0.88	0.99	0.76	0.81	0.78	18	22	19	14	33
Min		0.68	0.58	0.31	0.32	0.23	10	5	8	7	14

3.3.5 Discussion

Observer variability indicates the uncertainty in the experiment. For observer repeatability, mean and median R^2 values for all the observers were 0.69 and 0.72 respectively, and mean and median CV values were 18 and 17 respectively, as shown in Table 3-6. It is difficult to evaluate acceptability of the repeatability from this experiment; however, the observer repeatability obtained in a study to scale colourfulness (Luo and Clarke *et al.*, 1991) or gloss (Wei, 2006), for example, had CV values of around 19. According to the results of these previous experiments, the repeatability obtained in this experiment seems to be not too high or low. However, it should be noted that the observer variability are dependent on the attributes evaluated, the scaling methods and also background and experience of the observers. In comparison, observer accuracy was found to be reasonably high: the mean and median values of R^2 for all the observers were 0.82 and 0.83 respectively, and mean and median values of CV were 14 for both as shown in Table 3-7. It was found that both the repeatability and accuracy from the blue and purple colour samples showed poorer results compared to those of the green colour samples. This indicates that the scale values of the perceptual coarseness obtained from the green colour samples were more reliable than those of the blue and purple colour samples. Both the repeatability and accuracy were excellent for the grey colour samples in terms of R^2 (over 0.9). In contrast, the CV values were not as good as might be expected. These results indicate that the individual data are linearly related, but the absolute values of these data are not the same.

Parallel to the present study, the perceptual coarseness was investigated by Akzo Nobel using seven observers and 398 coating samples (Kirchner and Kieboom *et al.*, 2007). These samples were not only metallic-coating panels but also pearlescent-coating panels and some panels coated with a mixture of metallic and pearlescent pigments. They used a 0-9 categorical judgement scaling method with eight reference samples which corresponded to categories one to eight: category nine was used for any samples that were coarser than that indicated by category eight. More reference samples could be used in this experiment if there were appropriate samples available, for example the six grey colour samples could be used. However, the perceptual coarseness differences between these six samples and even the recipes of their coatings (the amount of aluminium flake contained in these coating samples) were unknown. Moreover, there was no method available to measure perceptual coarseness. Therefore,

only one reference sample was used. Another difference between their experiment and the present study was that they allowed the observers to assign the values up to a quarter scale precision. The visual assessment was carried out in their laboratory in which an isotropic and luminous flux was emitted all over the ceiling to provide diffuse illumination with the illumination and viewing geometry of about 0/45. The viewing condition used in their experiment is shown in Figure 3-7.

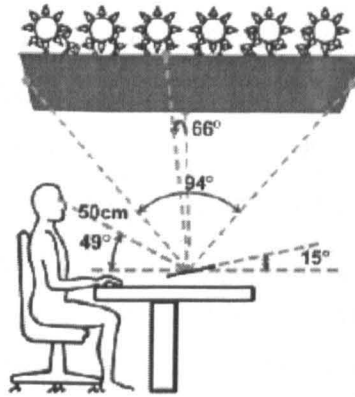


Figure 3-7: The viewing condition used for the coarseness assessment at Akzo Nobel (Kirchner and Kieboom *et al.*, 2007).

A mean value of observer accuracy reported was 0.90 in their study with minimum and maximum values of 0.83 and 0.92 respectively, in terms of the R^2 . These indicate better accuracy compared with the present study, where the mean, minimum and maximum values of the accuracy were 0.82, 0.68 and 0.88 respectively, in terms of R^2 . In their experiment, the observers “had experience in observing visual texture, and were trained to recognise visual texture attribute” (Kirchner and Kieboom *et al.*, 2007), while all the observers who participated in this study were not experienced in the visual assessment of visual texture. Although the target attribute is different, the professional observers consistently out-performed the student/naive observers regarding the ability to judge colour difference of metallic coatings (Nicholls, 2000) (see Section 2.5.1.1). In addition, they used eight reference samples compared with only one used in this study. More reference samples are likely to make the judgements easier and more precise. If these points are taken into consideration, it can be concluded that the accuracy obtained in the present experiment is not much worse than the result obtained using the professional observers. Hence, the relatively high observer accuracy in both experimental conditions encourages using the diffuse illumination for assessing the coarseness of the coatings and also indicates that the perceptual coarseness can be reliably assessed by observers.

3.4 Computational Model for Perceptual Coarseness Prediction

Image texture analysis has been much studied in the last few decades. A large number of image-texture analysis methods in the computing domain have been developed based on various analyses of image. However, those methods are mainly for *texture classification*, *texture segmentation* and *texture synthesis* (see Section 2.6) rather than for actually predicting visual texture using numerical values, although the performance of these tasks is often evaluated by visual inspection. Many of these methods require input parameters which are often difficult to determine (see Section 2.6). For example, although the comparison studies (Conners & Harlow, 1980; Ohanian & Dubes, 1992; Weszka and Dyer *et al.*, 1976) proved the ability of the *spatial grey level dependence matrix (co-occurrence matrix)* (see Section 2.6.1.1) for classification tasks, this method suffers from a number of difficulties because there are no well-established methods to select the directional and distance parameters.

The present author (Kitaguchi and Westland *et al.*, 2004) also implemented several conventional image texture analysis methods to investigate the correlation of features extracted from images of fabric samples with perceptual coarseness. The image texture analysis methods used and the features extracted from them are listed in Table 3-8.

Table 3-8: A list of image texture analysis methods implemented and features extracted in the coarseness assessment using the fabric samples.

Texture Analysis Method	Features
Spatial Grey Level Dependence Method (Co-occurrence matrices)	Energy Entropy Contrast
Run Length	Short run emphasis Long run emphasis Run length non-uniformity Grey level non-uniformity Run percentage
Grey Level Difference Method	Contrast Angular second moment Entropy Mean
Neighbouring Grey Level Dependence Statistics	Small number emphasis Number non-uniformity Large number emphasis Entropy Second moment

In the experiment, fifteen observers evaluated the coarseness of ten real 10×10 cm fabric samples in a viewing cabinet using the rank order method (Engeldrum, 2000). Each observer was asked to rank the samples in order of coarseness. In order to check their repeatability, each observer carried out the assessment twice. The raw experimental results were converted into z-scores. The mean values of the observers' z-score of coarseness with 95 % confidence intervals are given in Figure 3-8. The more negative z-scores indicate the perceptually coarser texture.

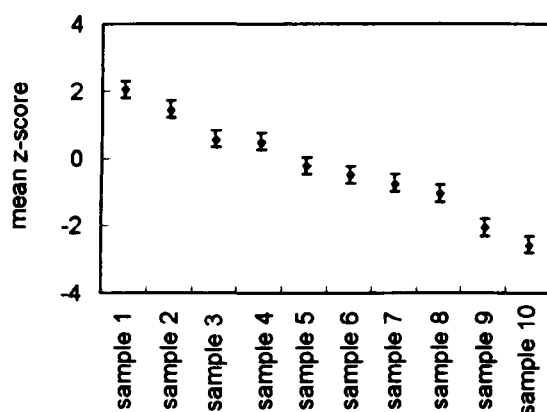


Figure 3-8: Mean of the observers' z-scores of coarseness with 95 % confidence intervals (the more negative z-scores indicate the perceptually coarser texture).

Among the features computed from the implemented methods, some, such as the contrast from the spatial grey level dependence method and the mean from the grey level difference method, showed high correlation with the perceptual coarseness, whose correlation, in terms of R^2 was 0.82 and 0.90 respectively. However, these features changed as the parameters varied. Although the feature describes the perceptual texture very well at a particular value of a parameter, the other parameters do not always show good correlation, as shown in Figure 3-9; the correlation for contrast from the spatial grey level dependence method changes when the distance parameter is changed from 1 to 11. In practice, it is necessary to consider the efficiency of the computation procedures. Although some of the methods showed good agreement with the perceptual coarseness, they were time consuming to examine and analyse because they involve many parameters. In addition, even human perception is affected by the viewing conditions such as the distance between an observer and the samples, and because of the size and resolution dependence of visual texture (see Section 1.2), the conventional methods are not so sensitive from the human perceptual perspective.

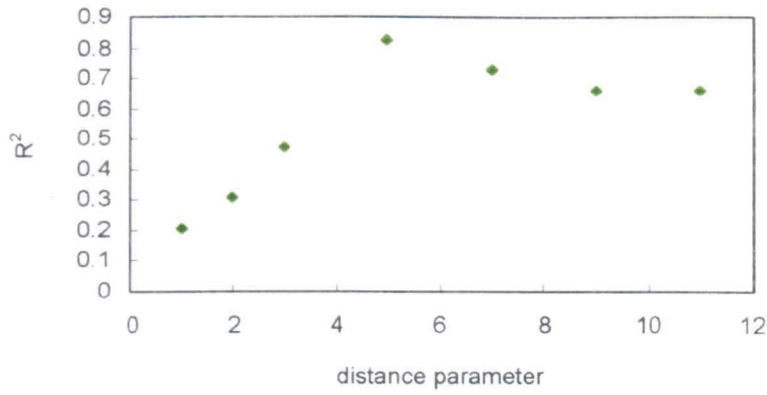


Figure 3-9: Changes of R^2 values for contrast from the spatial grey level dependence method when the distance parameter changes from 1 to 11.

The objective of the study described in this section is to develop a computational model capable of predicting the perceptual coarseness of the metallic coatings from a digital image of those coatings. Such a model should take into account the human visual system and analyse an image and it should also make a quantitative match with perceptual coarseness. Coarseness is caused by the non-uniformity of a surface and it can be analysed by looking at the pixel value variation in an image of that surface which is a measure of the contrast between neighbouring pixels. Therefore, a coarseness model was developed based on the hypothesis that the Fourier energy of an image is a measure of the amount of contrast in the image and that the amount of contrast is correlated closely with coarseness. Furthermore, the contrast-sensitivity function (CSF, see Section 2.2.3) measures the visible amount of the contrast which is related to perceptual coarseness. The model structure as given in Figure 3-10 is described in the following sections.

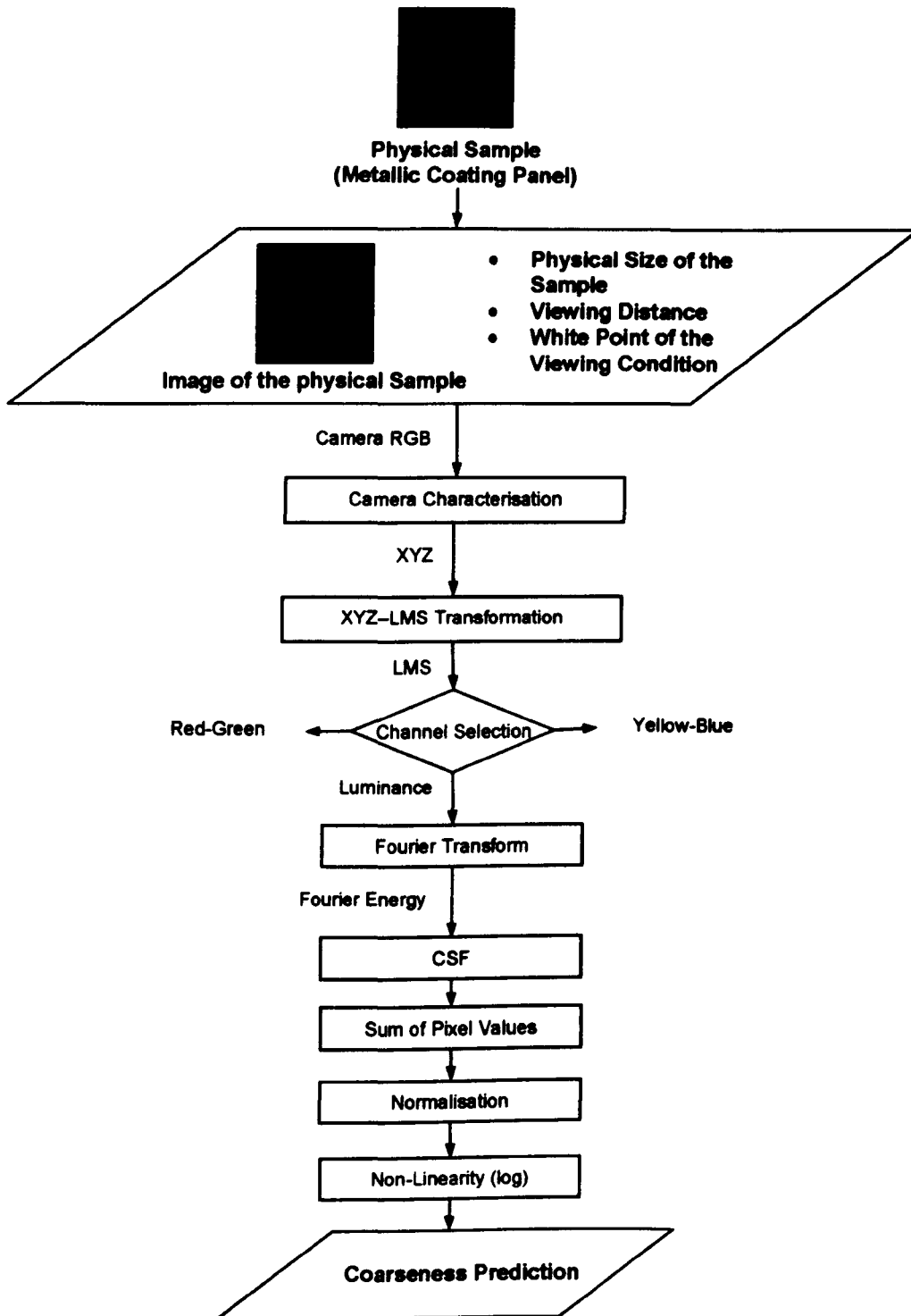


Figure 3-10: A flowchart of the main framework for the coarseness model.

3.4.1 Capture and Measurement of the Target

Spectroradiometers and spectrophotometers are often employed for the measurement of SPD and reflectance, parameters which can be associated with colour and gloss appearance respectively. The information obtained from these instruments is as a result of integration over a defined area. However, two-dimensional spatial

information is essential for analysis of visual texture (depending on the material, three dimensional information may be required). Therefore, a digital camera was used to obtain information on the spatial variation of a surface. In this section, the required data for a model and the development of a model are introduced.

Digital images and their spectral properties, such as the SPD and the CIE XYZ tristimulus values of the 156 metallic-coating samples (the same samples as used in the visual assessment in Section 3.3), 124 solid-colour-coating panels, a white ceramic standard tile and a GretagMacbeth ColorChecker DC (ColorChecker DC) were used in this experiment.

A Nikon D1X digital camera was used to capture images of all target samples: the metallic-coating samples, the solid-colour-coating panels, the white ceramic standard tile and the ColorChecker DC. This camera is a single-lens reflex (SLR) digital camera featuring a 23.7×15.6 mm CCD incorporated with a Bayer RGB filter, coded at 8-bits per channel and it can capture images of up to 3008×1960 pixels spatial resolution. Image capture was made with the same experimental conditions as used for the visual assessment (see Figure 3-3). A digital camera was placed at the approximately location of the observer's eyes as shown in Figure 3-11. For the metallic-coating samples and the solid-colour-coating panels, a part of the image (832×832 pixels), which corresponds to the sample size of 8×8 cm as used in the visual assessment, was selected from the whole image (3008×1960 pixels). These sub-images were selected not from the central area of the camera's view but at the location where the metallic-coating sample was placed in the visual assessment, so as to avoid the reflected mirror image.

The SPD of the metallic-coating samples, the solid-colour-coating panels and the white tile, were measured using the TSR. The TSR was mounted on the viewing cabinet, but it was tilted about 10° from the central area (Figure 3-11). This was again to avoid the reflected mirror image. The CIE XYZ tristimulus values were then calculated using the CIE 1964 standard colorimetric observer. Since the patches in the ColorChecker DC were too small to be measured in the cabinet, the spectral reflectance of all the patches was measured using the 7000A spectrophotometer (see Section 3.2.2). The CIE XYZ tristimulus values were then computed using the SPD of the illumination obtained by measuring the white ceramic standard tile and the CIE 1964 standard colorimetric observer data.

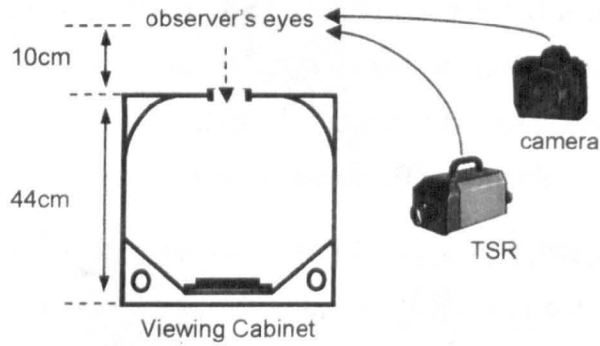


Figure 3-11: An illustration of the experimental condition for the measurement of the spectral properties using the TSR and the image capture using the digital camera.

3.4.2 Colour Space Transformation

3.4.2.1 XYZ Colour Space

A characterisation method (see Section 2.8.3) was employed in order to convert device-dependent image RGB values to device-independent values such as CIE XYZ tristimulus values.

In general, two data sets: a *training data set* and a *test data set*, are required. Both sets should consist of a set of device-dependent values and corresponding device-independent values, and the two data sets should be independent. A training data set is used to derive a model such as to generate the parameters of the transfer matrix that forms part of the characterisation model. The test data set is used to evaluate the performance of a model derived based on the training data set. Colour differences between measured and predicted XYZ values of the test data set indicate the performance of the derived model.

In this study, a variety of linear and polynomial regression models with least-squares fitting (see Section 2.8.3.1) were first implemented and then an appropriate model was selected, which transforms the colours from RGB to XYZ values with a minimum error. Size of the transfer matrix (M) and augmented matrices (D) (see Section 2.8.3.1) used in this study are given in Table 3-9.

Table 3-9: Size of the transfer matrix and augmented matrices compared in this study.

Size of Transfer Matrix (M)	Augmented Matrices (D)
3 × 3	[R G B]
3 × 4	[R G B 1]
3 × 5	[R G B RGB 1]
3 × 9	[R G B RG RB GB R ² G ² B ²]
3 × 10	[R G B RG RB GB R ² G ² B ² 1]
3 × 11	[R G B RG RB GB R ² G ² B ² RGB 1]
3 × 20	[R G B RG RB GB R ² G ² B ² RGB R ² G R ² B G ² R G ² B B ² R B ² G R ³ G ³ B ³ 1]
3 × 35	[R G B RG RB GB R ² G ² B ² RGB R ² G R ² B G ² R G ² B B ² R B ² G R ³ G ³ B ³ R ³ G R ³ B G ³ R G ³ B B ³ R B ³ G R ² GB RG ² B RGB ² R ² G ² R ² B ² G ² B ² R ⁴ G ⁴ B ⁴ 1]
3 × 56	[R G B RG RB GB R ² G ² B ² RGB R ² G R ² B G ² R G ² B B ² R B ² G R ³ G ³ B ³ R ³ G R ³ B G ³ R G ³ B B ³ R B ³ G R ² GB RG ² B RGB ² R ² G ² R ² B ² G ² B ² R ⁴ G ⁴ B ⁴ R ⁴ G R ⁴ B G ⁴ R G ⁴ B B ⁴ R B ⁴ G R ³ GB RG ³ B RGB ³ R ³ G ² R ³ B ² G ³ R ² G ³ B ² B ³ R ² B ³ G ² R ² G ² B R ² GB ² R G ² B ² R ⁵ G ⁵ B ⁵ 1]

The purpose of the characterisation in this experiment is to transform RGB values of the image of the metallic-coating samples to CIE XYZ values on a pixel-by-pixel basis. Hence, mean RGB values for each of the images of the metallic-coating samples and the XYZ values of the corresponding samples based on the measurements using the TSR were used as a test data.

Careful consideration must be given to the choice of colours for the training data set. Available data in this experiment were the mean RGB values for each image of the 156 metallic-coating samples, the 166 patches in the ColorChecker DC and the 124 solid-colour-coating panels and the measured XYZ values of the corresponding samples. The colour distribution of these data are given in a CIELAB L*C* diagram and a CIELAB a*b* diagram shown in Figure 3-12 (a) and (b) respectively.

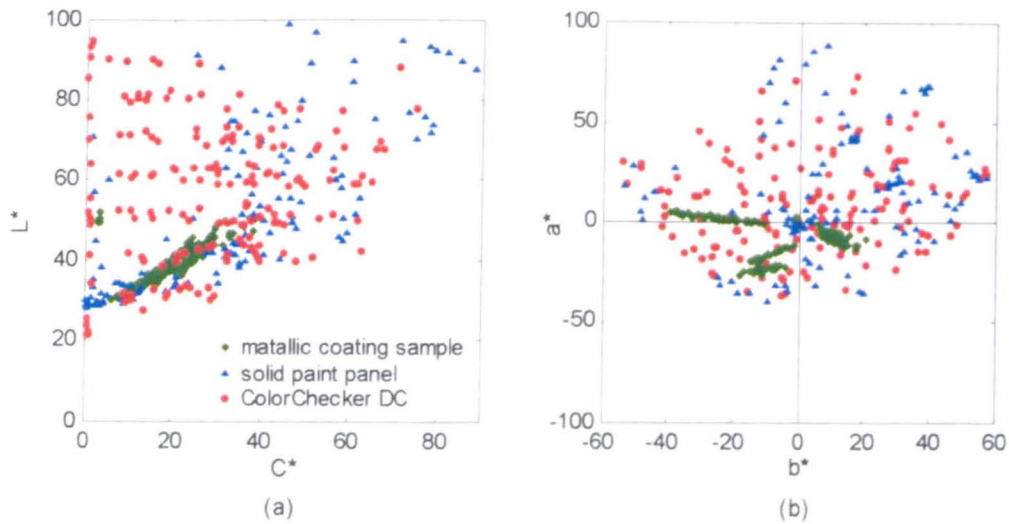


Figure 3-12: Colour distributions of the 156 metallic-coating samples, 124 solid-colour-coating panels and 166 patches in the ColorChecker DC. (a) A CIE LAB L^*C^* diagram. (b) A CIE LAB a^*b^* diagram.

The ColorChecker DC is often used as training data (see Section 2.8.3.1). Also, it is common to omit the patches that have a high gloss surface and only consider the matt surface patches which form the majority in it (Cheung, 2004). This is because that surface material differences cause errors to a characterisation model. This experiment deals with the metallic coatings whose surface finish differs from that of the ColorChecker DC. But, in practice, it is also not unusual for a chart being used as training data, even though its surface finish differs from that of the samples being considered. The solid-colour-coating panels did not have exactly the same surface material as the metallic-coating samples, but it was similar. At least, their glossy surface materials were closer to that of metallic-coating samples compared with the matt surface of the majority patches in the ColorChecker DC. Moreover, the solid-colour-coating panels included a wide range of colours, in contrast to the variation in the colours of the metallic-coating samples which was very limited. In order to find a best set of training data, consequently, four different combinations of these data were employed as training data sets as summarised in Table 3-10. The first training data set (Data Set 1) consisted of the 156 metallic-coating samples. The second data set (Data Set 2) included the 124 solid-colour-coating panels. The third data set (Data Set 3) was consisted of the 166 patches in the ColorChecker DC (excluding glossy surface patches). For the fourth data set (Data Set 4), the 124 solid-colour-coating panels and the 156 metallic-coating samples were categorised as training data. The models derived from each training data set were then evaluated using the test data set consisting of the 156 metallic-coating samples which were also included in Data Set 1 and Data Set 4.

Therefore, in case of Data Set 1, a leave-one-out method was applied so that data of 155 metallic-coating samples out of 156 were used as training data and a model was evaluated with data of the one remaining metallic-coating sample. Consequently, 156 trials were needed to be made in order to evaluate a model with all test data. Similarly, in case of Data Set 4, all the 124 solid-colour-coating panels and the 155 metallic-coating samples out of the 156 metallic-coating samples were used as training data and the remaining one metallic-coating sample was used to test the performance. Again, 156 trials were needed.

The summary of the data sets and the models that best performed is given in Table 3-10, and the performance of all the models derived from Data Set 1, Data Set 2, Data Set 3 and Data Set 4 is found in Table 3-11 to Table 3-14 respectively. The performance was evaluated in terms of CIELAB ΔE^*_{ab} between the measured and predicted values for both the training and test data sets.

It is shown in Table 3-11 to Table 3-14 that the smallest median test errors for Data Set 1, Data Set 2, Data Set 3 and Data Set 4 are ΔE^*_{ab} values of 0.36 for $M = 3 \times 35$, 1.39 for $M = 3 \times 35$, 1.49 for $M = 3 \times 35$ and 0.65 for $M = 3 \times 56$ respectively. The models derived from Data Set 1 performed best in significance level ($p < 0.05$) according to the Wilcoxon signed-rank test (see Section 2.10.5). The models obtained from Data Set 1 were generally better than those obtained from the other training data sets with any number of terms. Therefore, Data Set 1 was considered to provide the most appropriate model in terms of the test errors, and also in terms of the surface material. This is because the samples consisted of the metallic-coating samples themselves, so there were no surface material differences between the training and the test data sets. However, a shortcoming of Data Set 1 was that the colour distribution of the training data was not widely spread in colour space as seen in Figure 3-12. If the model would be applied to samples having much wider colour distribution than Data Set 1, then Data Sets 2, 3 or 4 which all include a wider range of colours would be more appropriate. However, in this experiment, the model was applied to the each pixel in the images of the metallic-coating samples and therefore the colour distribution of these pixel values was expected to be wider but not too different from the mean pixel values of each image of the metallic-coating samples, *i.e.*, Data Set 1. Consequently, Data Set 1 was chosen as a training data set for this experiment. The smallest median test error obtained was ΔE^*_{ab} of 0.36 from $M = 3 \times 35$ and an equivalently smaller error of 0.41 ΔE^*_{ab} was obtained from $M = 3 \times 20$. In contrast, the maximum test error of the model from $M =$

3×35 was ΔE^*_{ab} of 4.73 which was a much larger compared with that of 1.77 ΔE^*_{ab} obtained from $M = 3 \times 20$. This large maximum error for $M = 3 \times 35$ was likely caused by over-fitting the training data. Consequently, the polynomial model ($M = 3 \times 20$) derived from all the 156 samples was applied to the samples to transform the image RGB values to CIE XYZ values on a pixel-by-pixel basis.

Table 3-10: A summary of the data sets and the performance of the characterisation model which showed the smallest test error for each training data set.




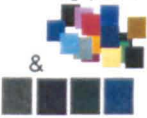
	Data Set 1	Data Set 2	Data Set 3	Data Set 4	
Training Data	155 metallic-coating samples 	166 patches in ColorChecker DC 	124 solid-colour-coating panels 	156 metallic-coating samples + 124 solid-colour-coating panels & 	
Test Data	156 metallic-coating samples	156 metallic-coating samples	156 metallic-coating samples	156 metallic-coating samples	
Method	Leave-one-out	–	–	Leave-one-out	
Size of M	3×20	3×35	3×35	3×56	
Test Error	Median	0.41	1.39	1.49	0.65
	Mean	0.51	1.54	1.66	0.73
ΔE^*_{ab}	Max	1.77	4.03	5.01	1.95
	Min	0.06	0.33	0.32	0.06

Table 3-11: Model performances using the Data Set 1.

Data	Training				Test			
	155 metallic-coating samples				156 metallic-coating samples			
Size of M	Training Error (ΔE^*_{ab})				Test Error (ΔE^*_{ab})			
	Median	Mean	Max	Min	Median	Mean	Max	Min
3×3	1.92	2.22	7.42	0.06	1.93	2.27	7.47	0.24
3×4	1.83	2.06	6.90	0.07	1.88	2.12	6.95	0.19
3×5	1.76	1.84	7.33	0.07	1.81	1.91	7.38	0.20
3×9	0.46	0.62	2.66	0.03	0.48	0.66	3.18	0.04
3×10	0.45	0.61	2.63	0.01	0.48	0.65	3.07	0.02
3×11	0.45	0.61	2.64	0.01	0.50	0.66	3.14	0.01
3×20	0.35	0.44	1.50	0.03	0.41	0.51	1.77	0.06
3×35	0.26	0.34	1.25	0.02	0.36	0.50	4.73	0.04
3×56	0.21	0.28	1.18	0.01	0.37	0.60	6.35	0.05

Table 3-12: Model performances using the Data Set 2.

	Training				Test			
Data	124 solid-colour-coating panels				156 metallic-coating samples			
Size of M	Training Error (ΔE^*_{ab})				Test Error (ΔE^*_{ab})			
	Median	Mean	Max	Min	Median	Mean	Max	Min
3 × 3	14.08	16.26	54.68	3.24	11.14	11.82	18.25	6.34
3 × 4	19.73	21.93	83.48	3.98	15.55	15.96	34.96	5.12
3 × 5	17.40	20.20	111.71	2.50	11.75	15.00	37.28	4.25
3 × 9	6.52	7.12	18.61	0.98	7.81	7.40	11.89	0.57
3 × 10	6.79	6.99	19.11	0.99	7.77	7.62	12.72	3.41
3 × 11	6.17	6.68	18.01	1.41	5.90	6.38	11.87	1.79
3 × 20	2.35	2.96	10.16	0.30	3.43	3.86	8.24	0.30
3 × 35	1.44	1.73	6.20	0.15	1.39	1.54	4.03	0.33
3 × 56	0.94	1.17	3.93	0.08	1.63	1.94	5.57	0.55

Table 3-13: Model performances using the Data Set 3.

	Training				Test			
Data	166 patches in the ColorChecker DC				156 metallic-coating samples			
Size of M	Training Error (ΔE^*_{ab})				Test Error (ΔE^*_{ab})			
	Median	Mean	Max	Min	Median	Mean	Max	Min
3 × 3	12.74	13.25	35.01	2.01	14.25	13.98	15.34	9.77
3 × 4	11.11	22.71	110.78	2.48	16.62	18.51	57.42	3.59
3 × 5	9.05	13.01	123.28	0.83	15.61	17.68	37.41	2.89
3 × 9	5.18	6.31	18.69	0.69	8.40	8.30	14.80	1.61
3 × 10	5.22	5.62	18.53	0.90	7.03	7.51	14.30	3.54
3 × 11	4.10	5.05	19.58	0.46	6.14	6.27	9.11	2.42
3 × 20	2.00	2.54	12.65	0.25	3.15	3.46	6.87	0.86
3 × 35	1.46	1.88	12.35	0.24	1.49	1.66	5.01	0.32
3 × 56	1.20	1.53	12.10	0.22	1.84	2.12	5.07	0.43

Table 3-14: Model performances using the Data Set 4.

	Training				Test			
Data	124 solid-colour-coating panels + 155 metallic-coating samples				156 metallic-coating samples			
Size of M	Training Error (ΔE^*_{ab})				Test Error (ΔE^*_{ab})			
	Median	Mean	Max	Min	Median	Mean	Max	Min
3 × 3	10.93	11.95	35.27	3.31	9.97	9.56	17.74	4.84
3 × 4	11.88	13.21	60.81	1.01	9.58	9.82	20.27	1.18
3 × 5	10.40	12.70	119.97	0.69	7.68	7.93	19.73	1.78
3 × 9	4.20	4.92	19.03	0.13	3.77	3.49	7.29	0.16
3 × 10	4.15	4.92	19.33	0.25	3.79	3.68	7.46	0.32
3 × 11	4.29	4.75	18.37	0.29	3.52	3.41	6.37	0.49
3 × 20	1.97	2.29	12.42	0.08	1.75	1.75	3.11	0.16
3 × 35	0.92	1.23	7.77	0.05	0.75	0.85	2.54	0.07
3 × 56	0.76	0.94	4.39	0.02	0.65	0.73	1.95	0.06

3.4.2.2 LMS Colour Space

The human visual system processes images in a way that is consistent with a spatial-frequency analysis of an image (Wandell, 1995). By taking this into consideration, it is assumed that the CSF should be incorporated into the model in order to accomplish a spatial-frequency analysis in terms of the human visual system. As introduced in Section 2.2.3, the contrast sensitivity of the human visual system is different in the luminance channel, the red-green channel and the yellow-blue channel. Thus the image of the metallic-coating samples had to be separated into these three channels. To do this, the XYZ values of each pixel in the image of the samples were first transformed into the LMS cone fundamentals (see Section 2.2.2.1). The following transformation as shown in Equation 3-3 was proposed by Stockman and MacLeod *et al.* (1993) enable transformation of the CIE 1964 XYZ CMFs ($\bar{x}_{10}(\lambda), \bar{y}_{10}(\lambda), \bar{z}_{10}(\lambda)$) to corresponding LMS cone fundamentals ($l(\lambda), m(\lambda), s(\lambda)$). However, the spectral property of each pixel in the image was not recovered in this study, since a polynomial regression model was employed to transform from the RGB to XYZ values. Instead, the simplified transformation between XYZ to LMS values as shown in Equation 3-4 was employed. The estimated S-cone response was not quite correct because there is no linear conversion equation; two different equations need to be applied at wavelengths either shorter than or longer than 520 nm. However, the effect was negligible in this study, since the S-cone information was not used in the final model.

$$\begin{aligned}
 l(\lambda) &= 0.236157 \bar{x}_{10}(\lambda) + 0.826427 \bar{y}_{10}(\lambda) - 0.045710 \bar{z}_{10}(\lambda) \\
 m(\lambda) &= -0.431117 \bar{x}_{10}(\lambda) + 1.206922 \bar{y}_{10}(\lambda) + 0.090020 \bar{z}_{10}(\lambda) \\
 s(\lambda) &= 0.040557 \bar{x}_{10}(\lambda) - 0.019683 \bar{y}_{10}(\lambda) + 0.486195 \bar{z}_{10}(\lambda) \\
 \log_{10} s(\lambda) &= 10402.1 / \lambda - 21.7185 \quad (\lambda > 520nm)
 \end{aligned}$$

Equation 3-3

$$\begin{aligned}
 L &= 0.236157 X + 0.826427 Y - 0.045710 Z \\
 M &= -0.431117 X + 1.206922 Y + 0.090020 Z \\
 S &= 0.040557 X - 0.019683 Y + 0.486195 Z
 \end{aligned}$$

Equation 3-4

The three channels; luminance, red-green and yellow-blue, were then separated from the image according to a chromaticity coordinate system proposed by MacLeod and Boynton (1979) (see Section 2.2.2.2). The luminance channel was determined from the values of L + M; the red-green channel from L / (L + M); the yellow-blue channel from S / (L + M).

3.4.3 Contrast Measure

A two-dimensional discrete Fourier transform (see Section 2.6.2.1) was applied to each of the three channels (luminance, red-green and yellow-blue) to transfer the spatial domain into the frequency domain to follow the assumption that the amount of energy in the Fourier transform of the image was a measure of the amount of contrast in that image. It should be noted that a uniform image by definition has zero coarseness. It also has zero contrast and zero energy apart from its DC component. The DC component signifies a direct current, the energy in the Fourier transform at the origin of the frequency domain; *i.e.*, zero frequency. As the brightness of an image increases, the DC component increases accordingly. Before applying the Fourier transform to the image, the mean value of each luminance, red-green and yellow-blue channel was subtracted from every pixel value in each channel, so that the DC component in the Fourier transform was zero.

Figure 3-13 shows an original image of a sample and its Fourier spectrum images for the three channels, where the top left image is the original image, the top right is a Fourier spectrum image for the luminance channel, the bottom left is the red-green channel and the bottom right is the yellow-blue channel. Note that the appearance of the images in this figure might be different from the physical sample because it was not possible to apply colour management to the images printed in this document. These images indicate that there is little Fourier energy in both chromatic channels and that there is large amount of Fourier energy in the luminance channel. This suggests that, for these samples, the chromatic channels make little contribution to perceptual coarseness. In addition, it is known that the spatial resolution of the human perception is more sensitive to luminance variation than chromatic variation (see Sections 2.2.2 and 2.2.3). Therefore, this study focused only on the luminance channel for modelling coarseness.

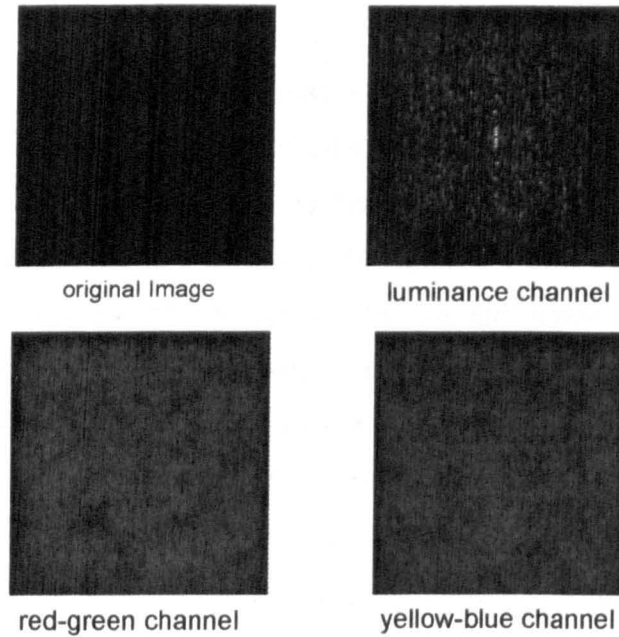


Figure 3-13: An original image of a metallic-coating panel (top left) and its Fourier spectrum images for luminance (top right) and chromatic channels: red-green channel (bottom left) and yellow-blue channel (bottom right). Note that the DC component is in the centre of each Fourier spectrum image and that the spatial frequency increases from the centre to outwards.

3.4.4 Applying CSF

As has been mentioned before, to incorporate properties of the human visual system into the model, the Fourier energy, which was the measure of the contrast contained in the image, was weighted using the CSF modelled by Westland (2005) on the basis of the CSF measurements made by Owens (2002) (see Section 2.2.3.2) as shown in Equation 3-5, and the sum of these weighted values was computed.

$$CSF(u) = F_L F_C \times 0.28u \exp(-0.3u)[1 + \exp(0.3u)]^{0.5}$$

where

$$F_L = \left\{ \begin{array}{ll} 1000(L/70)^{1/3} & \text{if } 1 \leq L \leq 70 \\ 1000(1/70)^{1/3} & \text{if } L < 1 \\ 1000 & \text{if } 70 < L \end{array} \right\} \quad \text{Equation 3-5}$$

$$F_C = (1 - d)$$

$$d = [(x - x_{white})^2 + (y - y_{white})^2]^{0.5}$$

where u is the spatial frequency in cycles/degree, L is the mean luminance level of a sample in units of cd/m^2 , d is a measure of the chromatic content of the image, (x, y) is the average chromaticity co-ordinate derived from the XYZ values for a sample image

and (x_{white}, y_{white}) is the chromaticity of the white point. Consequently, information about the viewing geometry of the experimental conditions provides input parameters to this model.

3.4.5 Normalisation

Comparison of the model output (the sum of the Fourier energy weighted using the CSF) with the scale values of the perceptual coarseness (results of the visual assessments described in Section 3.3.4.1) of the samples require that the data be normalised. The sum of the Fourier energy weighted using the CSF was normalised using the mean value of the luminance channel according to characteristic of the human contrast sensitivity, in which the ratio of the increment threshold to the background intensity is said to be a constant. This is explained with *Weber's law* (see Section 2.2.3.1). Figure 3-14 shows the original images of the grey colour samples on the middle row and their Fourier spectrum images in the luminance channel on the bottom row along with the scale values of the perceptual coarseness on the top row. It can be seen that the lightness values of these original images are similar, but the perceptual coarseness increases as the Fourier energy becomes larger. When comparing dark colour samples with light colour samples of similar perceptual coarseness, it was found that the Fourier energy was much greater for the light samples than for dark samples, as shown in Figure 3-15 which shows the original images of the green colour samples on the middle row and their Fourier spectrum images on the bottom row along with the coarseness scale values on the top row. Figure 3-16 also demonstrates this finding by plotting the relationship between Fourier energy weighted by the CSF and the mean value of the luminance channel of the images of the samples having similar perceptual coarseness. The four points in Figure 3-16 correspond to the same four sample images as is shown in Figure 3-15. According to the coarseness scale values, these samples are expected to be similar in appearance. However, it is evident that, as shown in Figure 3-16, the sum of the Fourier energy weighted by the CSF is much greater for the light samples. Therefore, this effect was included in the model by normalising such a sum using the mean value of the luminance channel for each image.

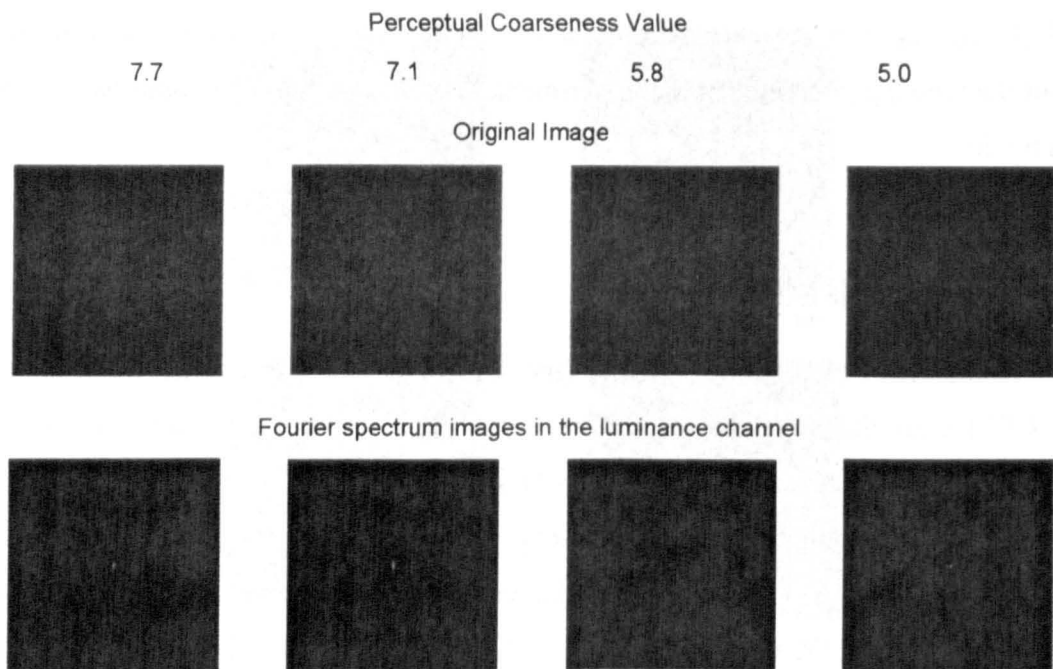


Figure 3-14: Perceptual coarseness values obtained from the visual assessment of grey samples (top row). Original images of the samples (middle row). Fourier spectrum images of the luminance channel (bottom row).

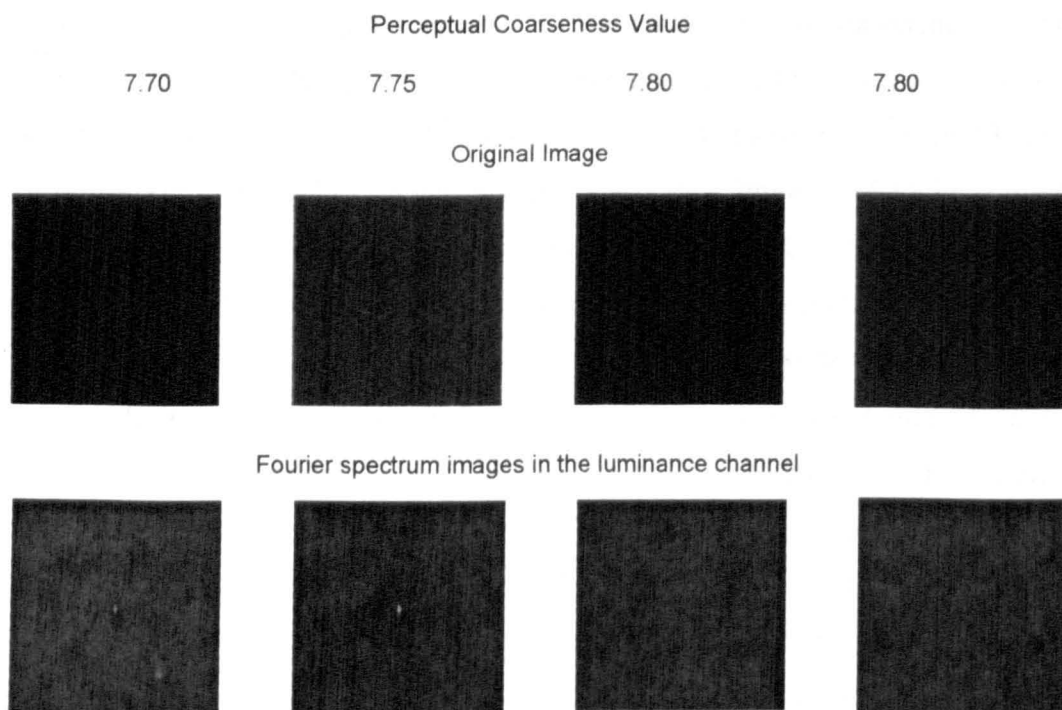


Figure 3-15: Perceptual coarseness values obtained from the visual assessment of green samples (top row). Original images of the samples (middle row). Fourier spectrum images of the luminance channel (bottom row).

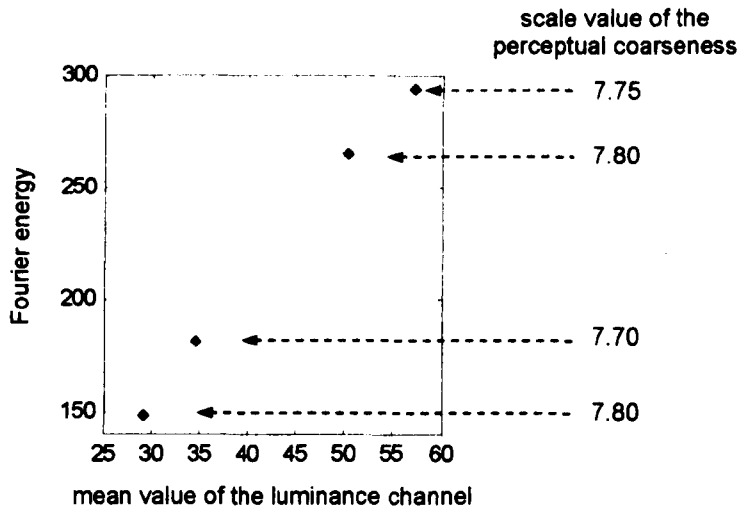


Figure 3-16: An example of the relationship between Fourier energy and the lightness of images in the green colour samples which are the same samples as in Figure 3-15 having similar perceptual coarseness (a scale value of 7.70 – 7.80) of the perceptual coarseness obtained from the visual assessment.

3.4.6 Non-Linearity

The output values after the process introduced in the previous sections are compared by plotting against the scale values of the perceptual coarseness for all 155 samples in Figure 3-17. It appears that the perceptual coarseness tends to increase approximately linearly with the logarithmic of the output values. This is not an unexpected finding. It has been found that human responses do not usually have a linear relationship with any associated physical properties, which has been evidenced from many psychophysical studies (see Section 2.9.2). Therefore, a function was derived by comparing the outputs with the coarseness scale values and a logarithmic function was found to be appropriate to describe the nonlinear behaviour of the output values. Incorporating this, the model can be express using Equation 3-6.

$$CoarsenessModel = \log_{10} \left(\frac{\sum_0^{u_{max}} E(u) \times CSF(u)}{L \times S} \right) \quad \text{Equation 3-6}$$

where u is the spatial frequency in cycles/degree, u_{max} is the maximum spatial frequency containing in an image, $CSF(u)$ is the CSF given in Equation 3-5, $E(u)$ is the Fourier energy, L is the mean value of the luminance channel and S is the size of an image in pixel units. The model prediction for each metallic-coating sample at the experimental condition as described in Section 3.3.2.1 is given in Appendix II.

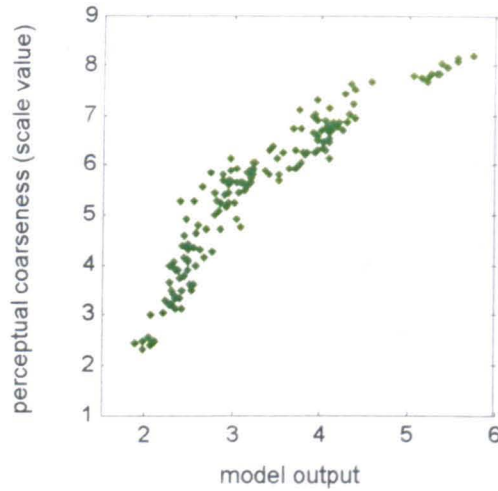


Figure 3-17: Comparison between the output values before linearisation and the scale values of the perceptual coarseness values.

3.5 Model Performance

The model developed in Section 3.4 was evaluated by comparing its output with the scale values of the perceptual coarseness obtained from the visual assessment using the metallic-coating panels described in Section 3.3. It should be noted that the model was developed on the image statistics and vision model which were completely independent from the scale values of the perceptual coarseness, except the non-linearity stage.

The model predictions and the scale values for all the 155 metallic-coating panels are plotted in Figure 3-18, and those of the grey, blue, green and purple colour samples are individually shown in Figure 3-19, with the accuracy of the model predictions investigated in terms of the coefficient of determination (R^2) and the coefficient of variation (CV) (see Section 2.101 and 2.10.3). Since the model predictions and the scale values do not have the same units, scaling factors (SF) were computed so as to correct their units for obtaining CV.

The result of an R^2 value of 0.91 for all the samples suggests an excellent relationship between the model prediction and the perceptual coarseness and indicates better performance than the observer repeatability (0.72) and accuracy (0.83). R^2 values for the grey, blue, green and purple colour samples were 0.96, 0.79, 0.95 and 0.81 respectively. Slightly lower model accuracy was found for the blue and purple colour samples. According to the result of the visual assessments, both the observer

repeatability and the accuracy of these two colour samples were also poorer than the other colour samples, as has been seen in Table 3-6 and Table 3-7 (see Section 3.3.4.2). Therefore, the lower accuracy of the model performance for these samples might be caused by the reliability of the scale values rather than a fault with the model. In contrast, reliable scale values and accurate model prediction were both obtained for the green colour samples.

The CV values for all samples and for the grey, blue, green and purple colour samples were 7, 13, 6, 3 and 12 respectively. The CV values agreed well with the R^2 values, except the result for the grey colour samples. While an R^2 value is a measure of a linear relationship, a CV value measures an absolute difference (see Section 2.10.1 and 2.10.3). Although a scaling factor was applied in this analysis because of scale difference between the model predictions and the perceptual coarseness scale values, the poor CV value indicates that it was not possible to adjust this data set with a simple scaling factor. However, it is difficult to identify the cause of error due to the relatively small number of grey colour samples.

The model developed was only applied to the luminance channel of the image. Although the samples were coloured, the chromatic channels were not concerned. This was based on the fact that the Fourier energy of the chromatic channels in the image contained much less information than that in the luminance channel. Moreover, the human visual system is much more sensitive to the luminance than to the chromatic channels. The excellent model performance obtained for all the samples supports this theory. However, slight disagreement was observed between colours. It can be seen from Figure 3-18 that the model slightly under estimated for the blue colour samples compared with that for the green and purple samples or that the model slightly over predicted for green and purple samples. However, as shown in Figure 3-20, there were no significant differences for most of those samples, since the data were within 95 % confidence intervals. There were only a few blue colour samples that were significantly different from the others at a perceptual coarseness scale values around five. Hence, it was concluded that the influence of the chromatic channels can be considered negligible in this sample group. However, as has been shown in Figure 3-12, the colour distribution of this metallic-coating sample set used in this experiment was limited in colour space. Moreover, the perceptual coarseness variation did not quite overlap with the other colour sets of the samples. Therefore, in order to investigate a possible

chromaticity effect, it is necessary to carry out further experiments with more coloured samples. Further evaluations of the model are given in the following Chapter 4.

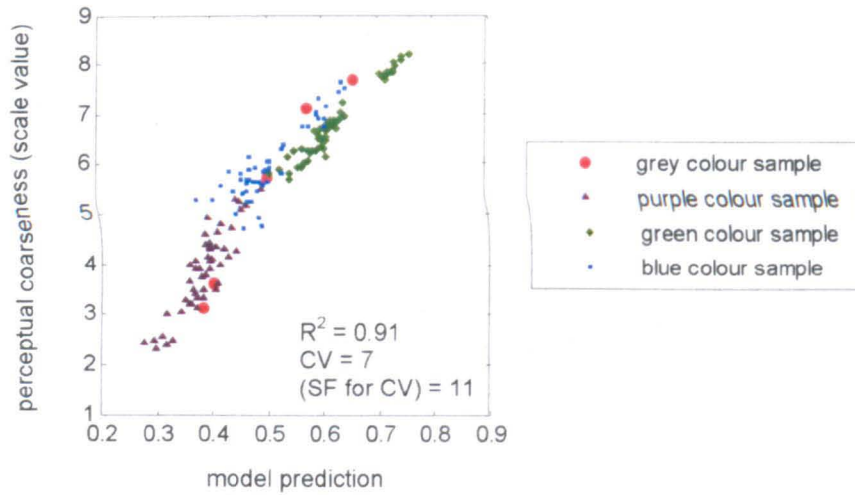


Figure 3-18: Comparison of the model predictions with the scale values of the perceptual coarseness for all the 155 metallic-coating panels.

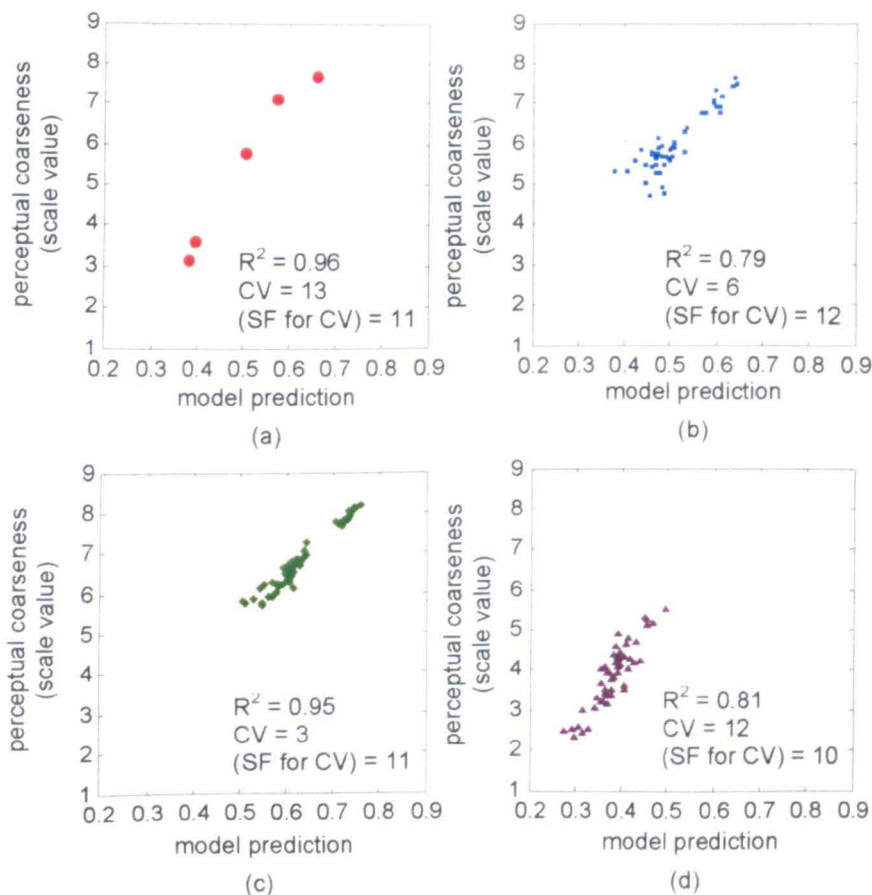


Figure 3-19: Comparison of the model predictions with the scale values of the perceptual coarseness for each colour metallic-coating panels: (a) grey colour samples, (b) blue colour samples, (c) green colour samples and (d) purple colour samples.

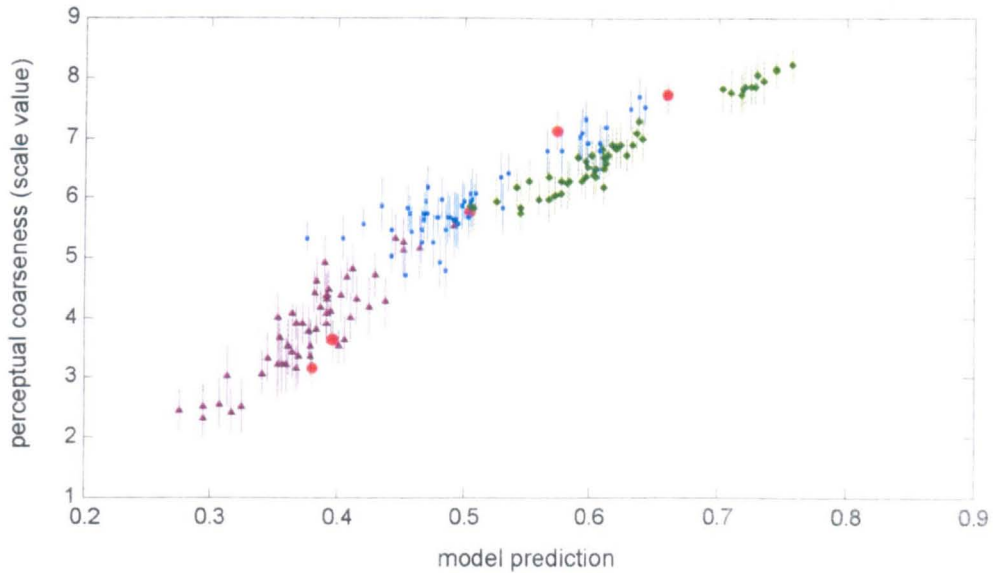


Figure 3-20: Comparison of the model predictions with the scale values of the perceptual coarseness for all the 155 metallic-coating panels with 95 % confidence intervals.

3.6 Summary

The characterisation of perceptual coarseness was carried out using a set of metallic-coating samples provided by Akzo Nobel. Because of gonioapparent nature of metallic coatings, the illumination and viewing conditions were carefully chosen so that the observers could consistently assess their coarseness. It was found that the coarseness could best be assessed under diffuse illumination. Using a 1-9 categorical judgement scaling method with one reference sample, the coarseness of the samples was visually scaled. The observer accuracy obtained from the visual assessment was less accurate in comparison with the coarseness assessment carried out in the similar conditions at Akzo Nobel (see Section 3.3.5). This difference perhaps originated from the different observers involved. While naive observers (had no experience in assessing the visual texture) participated in the present study, professional observers (had experiences in assessing visual texture) carried out the assessment at Akzo Nobel. Also, a 0-9 categorical judgement scaling method with eight reference samples was employed by Akzo Nobel. It could be thought that the use of more reference samples made the observers' judgement easier and it suggests using more than one reference sample to improve observer accuracy. In the present study, there were some variations in the observer accuracy depending on the colours of the samples. However, the high values of the overall observer accuracy and repeatability suggest the adequacy of the diffuse

illumination for the coarseness assessments and that the perceptual coarseness can be reliably assessed by observers.

A computational model was developed based on the information from a digital image of the metallic-coating samples. The advantages of this model are that it is computationally inexpensive and the input parameters (such as the physical size of the panel, the viewing distance and the white point for the observation condition) are related to the viewing conditions. Unlike the conventional texture models, such as SGLDM, GLDM, NGLDM and grey level run length (see Section 2.6), no parameters needed to be estimated to implement this model.

As it was for the visual assessment, suitable experimental conditions were critical for image capture and measurement of spectral properties which were necessary as model input. Therefore, they were carried out at the same conditions as the observers assessed the coarseness of the samples. The CIE XYZ tristimulus values corresponding to each pixel in the images were transformed using a camera characterisation. As a result of comparison of various training data sets that were used to derive such a model, the training data consisting of metallic-coating samples themselves were found to be the best. An advantage was that this data set did not have any surface material difference from the samples that the derived model was applied to. The XYZ values were then transformed to the LMS values and then to a luminance and two chromatic channels. The Fourier energy in these individual channels was computed to measure the amount of the contrast based on the assumption that the contrast correlated with coarseness. As a result, the large amount of the Fourier energy was contained in the luminance channel but less in the chromatic channels, so that only the luminance channel was utilised for next processes. The amount of the visible contrast was then measured by applying the CSF. The sum of the Fourier energy weighted using the CSF was then normalised by the mean value of the luminance channel for each image in order to incorporate Weber's law relating to contrast sensitivity. Finally, the coarseness prediction was obtained by taking the logarithm of such a sum after normalisation by the number of the pixels in the image.

The model performance was evaluated in comparison of the model predictions with the scaled perceptual coarseness of the samples. There were slight variations in the performance depending on the colours of the samples, for example, the model performance for the blue and purple colour samples was worse than that of the green. This might be influenced by observer variability which was slightly lower for the blue

and even lower for purple samples than the green. Although overall agreement between the model predictions and the scale values of the coarseness indicates that the luminance channel alone is sufficient for the prediction of coarseness, it would be interesting to carry out a further study to verify the model for use with a wider range of coarseness levels and colours of coating panels.

Chapter 4

Assessing and Modelling Coarseness using a Display

4.1 Introduction

In the development of the coarseness model, as described in Section 3.4, it was necessary to obtain coarseness information from metallic-coating samples via the images captured using a digital camera. A Nikon D1X camera was used to capture a physical area of $96 \times 96 \mu\text{m}$ on each pixel via a 5.4 megapixel CCD sensor. It is known however, that the physical size of the aluminium flakes, which are the major cause of the non-uniform appearance, or coarseness, of the metallic coatings had an approximate diameter of between 5 to 50 μm . This means that the resolution of the camera was not high enough to capture individual aluminium flakes in the metallic coatings. This can be overcome by taking images with higher resolution so as to capture a smaller area per pixel, but not sufficiently fine to cover the range of all the flakes using the camera used. However, it may not be necessary to capture higher resolution images, since the resolution of the human eye is also limited by its contrast-sensitivity function (CSF) (see Section 2.2.3). In addition, it is known that the perceptual attributes of metallic coatings are not only caused by the aluminium flakes. Observers see the reflected light not only directly from the flakes but also from other components such as colour pigments and the varnish clear coat, and the apparent size of the flakes is likely to be larger than their physical size because of the relationships between the incident light and its reflection (see Sections 2.4.1 and 2.5). Also, the reflection from diffuse illumination (which was used for coarseness assessments in Section 3.3) behaves more complicated manner than a directional incident light. Since no information about the structural composition of the metallic-coating samples, such as the paint recipes, the actual particle size distribution and the orientation of flakes in the samples, is available, it is not possible to estimate the reflected or scattered light from the surface of the samples in this experiment. Although it is possible to use a microscope to measure the reflected or scattered light of an area as small as a single flake (Sung and Nadel *et al.*, 2002), the requirement of this study is to analyse the appearance of metallic coatings rather than a single flake and hence a digital camera is most easily used to capture the spatial information in the form of an image; it is less costly and less time consuming than capturing micro-information with a microscope.

Investigations in this chapter were made in order to verify the reliability of the information in the image capture process described in Section 3.4.1, especially whether adequate information relevant to the perceptual coarseness was captured by the camera,

bearing in mind that the pixel size was larger than the flake size. Part of the verification process involved simply looking at the images to see if the coarseness appeared to match that of the original metallic-coating sample. If this can be achieved, an image capture system with higher resolution is then unnecessary. It also suggests that digital images of coating samples could be used to provide the stimuli for coarseness experiment rather using real samples. This would provide the opportunity, for example, to change the colour of a sample to produce a greater range of stimuli. It would also provide an extremely convenient facility for the product and quality control of coating systems, as well as a useful tool for product design and manufacturing.

Therefore, visual assessments were carried out to assess the coarseness using the images displayed on a liquid-crystal display (LCD) (see Section 2.8.2). The scaled coarseness of the images was compared with the perceptual coarseness of the actual metallic-coating samples described in Chapter 3.

To accomplish this, a set of images was assembled from those captured for the coarseness prediction described in the previous chapter (Section 3.4.1). The following sections start with the evaluation and characterisation of a display in order to reproduce accurate images on an LCD for this experiment. In practice, since it not possible to avoid the introduction of errors during an image generation process, these potential errors were analysed and are discussed. The visual assessments are then described including the viewing conditions and the actual coarseness scaling experiments. The image quality was then evaluated by comparing the perceptual coarseness results obtained using the images with that obtained using the actual metallic-coating samples. In addition, the coarseness model developed in Chapter 3 (see Section 3.4) was applied to the generated images. The predicted coarseness results were compared with the perceptual coarseness of the image to test the model's performance.

4.2 Display Evaluation

An Eizo ColorEdge CG220 LCD was used to display the images. This LCD had a size of 22.2 inch (56.4 cm) in diagonal and 1920×1200 pixel resolution with 8 bits per pixel. The chromaticity, the white point luminance and the gamma of the display were set to illuminant D65, 100 cd/m^2 and 2.2 respectively. This LCD can achieve luminance levels up to 200 cd/m^2 . In most situations however, this high level of luminance could cause fatigue in typical applications. Also, a wide luminance range

using only 8 bits colours would produce quantisation errors. Thus, a maximum luminance of 100 cd/m^2 was selected for this study. Measurements of the spectral properties for display evaluation and characterisation were made using a Minolta CS1000 spectroradiometer (TSR) (see Section 3.2.1) with the display in a darkened room. The TSR was always placed at a distance of 70 cm from the display. The measured targets were square with a size equivalent to 14 % of the full screen size (570×570 pixel) and were displayed in the centre area of the display with a mid-grey background (CIELAB L^* of approximately 50). It should be noted that influences of polarisation to both of the TSR and the display were not considered in this experiment. CIE XYZ tristimulus values were determined using the CIE 1964 standard colorimetric observer. Colorimetric errors were evaluated in terms of CIELAB ΔE^*_{ab} . The spectral power distribution (SPD) of the display primaries and the gamut of the display are plotted in Figure 4-1. As a reference, a gamut of sRGB colour space is also shown (IEC, 1998); note that the given sRGB gamut has been transformed to the values corresponding to the CIE 1964 standard colorimetric observer (Li, 2008). Different LCDs can have different gamuts but most have gamuts that are not too different from that of defined by the sRGB. It can be seen however, that this display has a particularly wide gamut in an area of the green primary but not in an area of blue, comparing with the sRGB gamut.

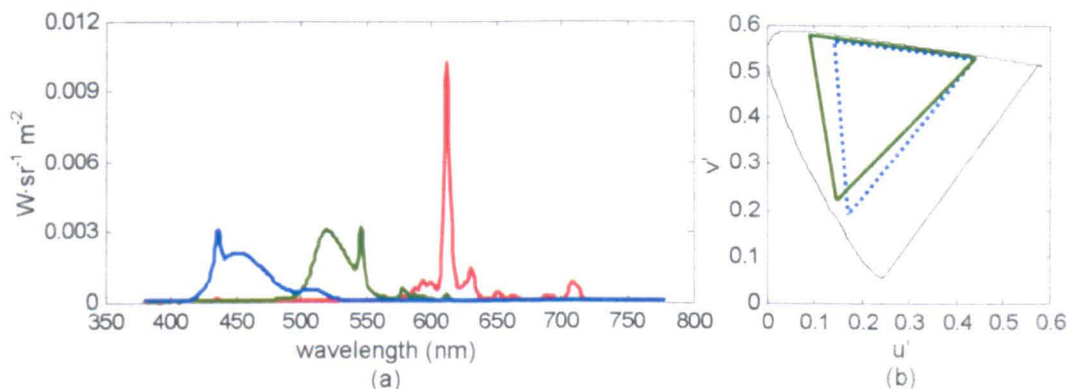


Figure 4-1: (a) SPD of the primaries of the LCD used in this experiment. (b) Chromaticity coordinates of the primaries of the LCD (solid line) and sRGB (dashed line) in the CIE 1976 Uniform Chromaticity Scale diagram.

4.2.1 Temporal Suitability

All displays require a finite time to reach a steady state from a cold start. A mid-grey colour patch was measured with the TSR each minute over the time period of 140 minutes. Temporal suitability was evaluated in terms of the colour difference, CIELAB

ΔE^*_{ab} , of each measurement from the last measurement (at 140 mins). The values of ΔE^*_{ab} were plotted against time as shown in Figure 4-2. It can be seen that there is a relatively large colour difference at the beginning; the value then decreased steadily after a period of approximately 15 minutes. These results indicate that the display needs to stabilise for at least 60 minutes to get a colour difference of less than 0.2 and for 100 minutes to reduce this figure to 0.1.

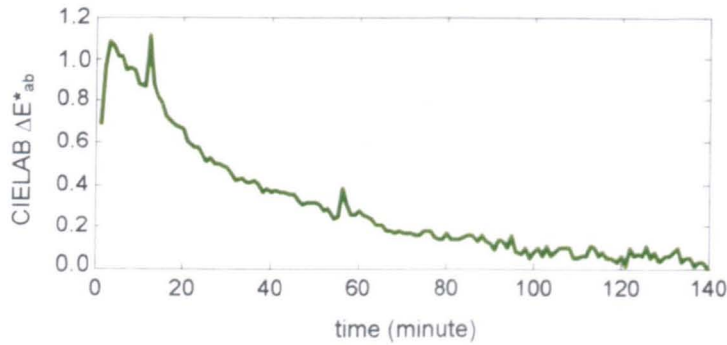


Figure 4-2: Temporal suitability of the display expressed in CIELAB ΔE^*_{ab} from the last measurement.

4.2.2 Additivity Test

Channel additivity of the display characterises the ability of that display to reproduce a colour which is the exact sum of the colours of the three primaries that comprise it. In order to evaluate this additivity, SPD measurements of each primary, red (R), green (G) and blue (G), were taken at each of thirty-three levels that corresponded to an approximately equal interval scale from 0 to 255, and XYZ values were then calculated from each measurement. The colours of the mixture of those primaries (R + G + B) were also measured at the same thirty-three levels and XYZ values calculated. The XYZ values of the black (R = 0, G = 0, B = 0) was subtracted from the XYZ values for each measurement. CIELAB parameters L^* , a^* and b^* were then calculated. Additivity errors were investigated by comparing these CIELAB values of both the colours from the separate primaries and from the sum of the three primaries. Figure 4-3 shows values of ΔE^*_{ab} , ΔL^* , Δa^* and Δb^* at each of the thirty-three levels where the measurements were made. Mean and maximum additivity errors were ΔE^*_{ab} of 1.17 and 1.61 respectively. It can be seen from Figure 4-3 that the errors of ΔL^* , Δa^* and Δb^* were not consistent in the range from 0 to 255. ΔL^* was larger in the middle of the range, but the Δa^* and Δb^* values were relatively larger at the high end of the range, although ΔE^*_{ab} was relatively consistent. At the peak output of the primaries (R = G = B = 255), the additivity errors were 0.55 %, 0.79 % and 0.58 % for

X, Y and Z values respectively. Gibson and Fairchild (2000) similarly evaluated additivity of two LCDs and found additivity errors of 0.01 %, 0.08 % and 0.09 % from one LCD and 7.61 %, 7.51 % and 2.98 % for the other for X, Y and Z values respectively. It seems that additivity can vary considerably between displays. The LCD used in this experiment exhibited larger errors than one of the displays in the study by Gibson and Fairchild, but smaller errors than the other. These errors suggests that it is necessary to take into account the influence of channel additivity (primary crosstalk) (Wen & Wu, 2006) to characterise this display.

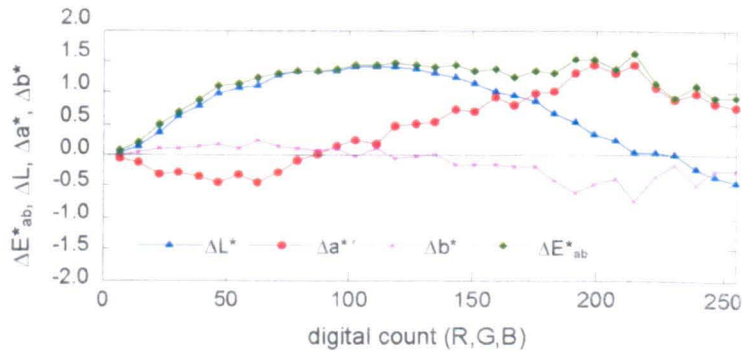


Figure 4-3: Additivity errors in terms of ΔE^*_{ab} , ΔL^* , Δa^* and Δb^* : For example, ΔL^* = (the separate measurements of the mixture of the three primaries) – (the addition of the measurements the three primaries).

4.2.3 Chromaticity Constancy of Primaries

Chromaticity constancy evaluates chromaticity changes of the primaries corresponding to the input digital counts of each channel. Thirty-three steps for each of the red, green and blue channels and grey (the additive mixture of three primaries) were measured as above. Chromaticity coordinates were calculated and the results plotted in the CIE 1976 Uniform Chromaticity Scales diagram as shown in Figure 4-4 (a). The chromaticity of all the colours follows a line from the point representing the primary towards the black (neutral). Ideally, there should be four individual points in Figure 4-4 (a) corresponding to each primary and grey, assuming there is no variation in the chromaticity coordinates. However, this is not found, probably due to the internal and external flare in the display. Therefore, it was necessary to re-calculate the chromaticity by subtracting the black from each measurement and these results are shown in Figure 4-4 (b). This black colour expresses the amount of internal and external flare, which should be removed. It can be seen that there are now clearly four points. However, some residual errors were still found. The green was very stable but changes were seen

for the red and blue channels. One reason for the chromaticity inconstancy originates in the physical nature of LCDs in that the spectral transmittance of liquid crystals varies as a function of the applied voltage and this causes the peak wavelength to shift toward shorter wavelengths with decreasing transmittance (see Section 2.8.3.3).

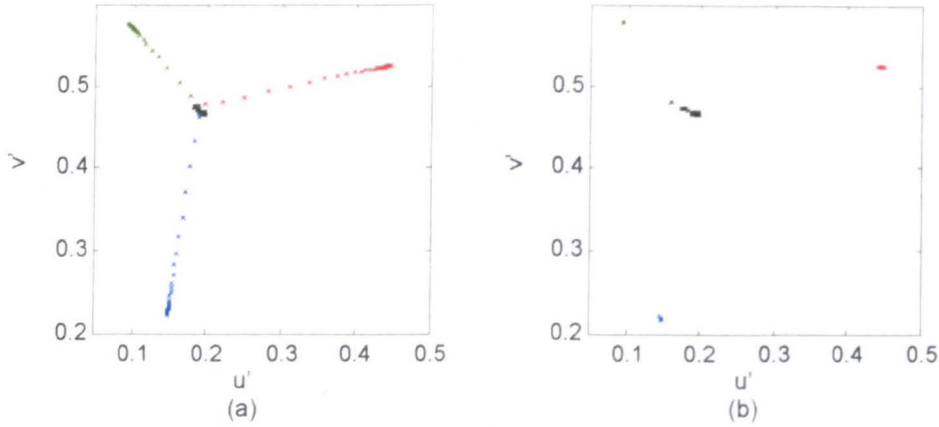
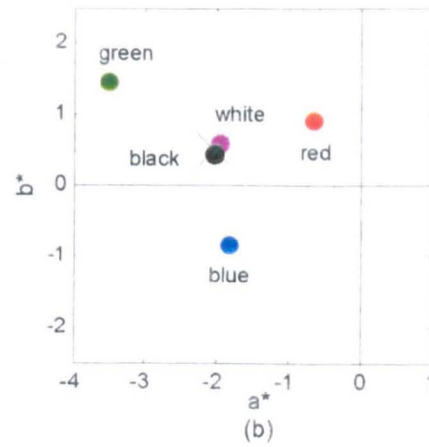


Figure 4-4: Chromaticity changes of each channel in CIE 1976 Uniform Chromaticity Scales diagram: (a) before; and (b) after black correction.

4.2.4 Spatial Independency

The output colour from one spatial location should not be affected by the colour of its surround. However, in practice, this spatial independency is not perfect. This was evaluated by measuring the colour differences between a grey patch displayed in the centre of the display with individual white, black, red, green and blue surrounds, and the physically identical grey patch with a grey surround. Values of ΔE^*_{ab} for each background colour are given in Figure 4-5 (a) and in addition the colour shifts are plotted in a CIELAB a^*b^* diagram in Figure 4-5 (b). The results clearly indicate that the colour of the central patch, the grey colour (indicated by cross in Figure 4-5 (b)), shifts toward the colour of the background by up to approximately 1.8 CIELAB ΔE^*_{ab} units.

Colour	Δ^*E_{ab}
White	0.82
Black	0.20
Red	1.46
Green	1.79
Blue	1.37



(a)

(b)

Figure 4-5: (a) Values of CIELAB ΔE^*_{ab} and (b) CIEALB a^*b^* parameters of the grey patch displayed at the centre surrounded by the same grey background (x) and by each of white, black, red, green and blue backgrounds (filled circles).

4.2.5 Spatial Uniformity

Ideally, colours should be reproduced uniformly across an entire display area. In order to evaluate the spatial uniformity of the display, measurements of white and grey patches were taken at nine different spatial locations arranged as a 3×3 grid pattern. The TSR was always placed normal to the screen. The colour differences between the colours measured at the centre and at each location are given as a representation of the spatial uniformity as shown in Figure 4-6. Also, the differences in value of CIELAB L^* at each location compared with the centre are given as a percentage in Figure 4-7. It is well known that most displays produce non-uniformity because of technological limitations. The display used in this experiment also showed some non-uniformity. The mean value of colour differences of the nine locations for the white and grey were very similar; ΔE^*_{ab} values of 2.44 and 2.47 respectively. In both cases, there were relatively large differences between at the centre and corner areas. In general, the luminance level reduces from the centre to the edges of most CRT displays (Berns, 1996). However, L^* increased toward the corners in case of this display (Figure 4-7). The cause of this slight increase is not known and although this non-uniformity was visible, no correction was applied.

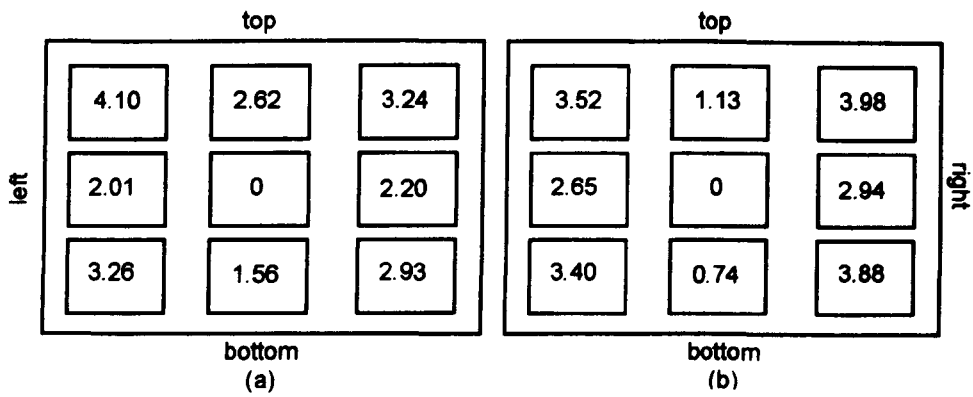


Figure 4-6: CIELAB ΔE^*_{ab} colour differences between the centre and nine locations evaluated for (a) white and (b) grey.

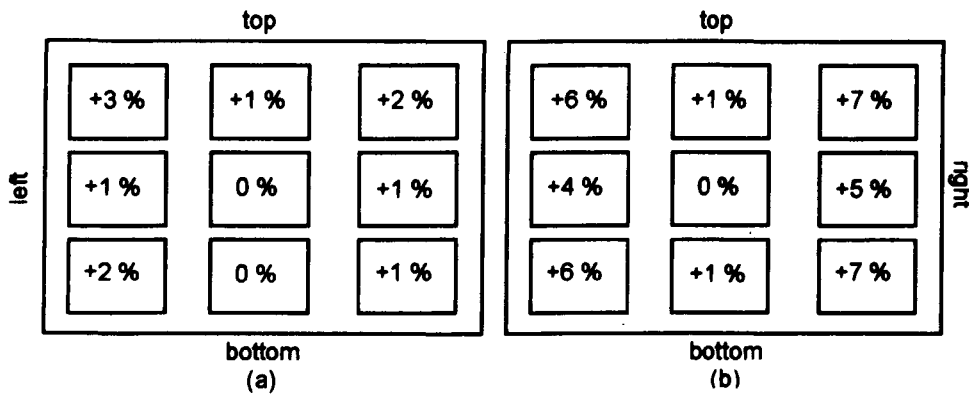


Figure 4-7: Percentage CIELAB L^* lightness differences between the centre and nine locations evaluated for (a) white and (b) grey.

4.3 Display Characterisation

Similar to camera characterisation, display characterisation defines a relationship between device-dependent values such as RGB values and device-independent values such as CIE XYZ tristimulus values (see Section 2.8.3). The technology for CRT displays has matured and a significant amount of research has been made into colorimetric characterisation (*e.g.*, Berns and Grozynski *et al.*, 1993; Berns and Motta *et al.*, 1993) (also see Section 2.8.3). For the characterisation of CRT displays, the gain-offset-gamma model, often referred to as the GOG model (see Section 2.8.3.2), is a well established method which takes into account the physical characteristics of CRT displays. Unlike CRT displays, however, the technology for LCDs is still immature and hence improving. This makes it difficult to develop a standard physical-based characterisation model. Therefore, several models were evaluated to select one which could best describe the characteristics of the LCD used in this experiment. Recent

LCDs tend to mimic the relationship (termed optoelectronic transfer function, OETF) between the input signals and the radiant power output of CRT displays and therefore the GOG model although developed for CRT displays, polynomial regression models (see Section 2.8.3.1), and the Day, Taplin and Berns's model (see Section 2.8.3.3) were implemented and then the model that provided the best performance was employed to obtain RGB values in order to best display the images of the metallic-coating panels on the LCD.

4.3.1 Data Sets

Three data sets consisting of RGB values and XYZ values were created to derive and test the characterisation models. The first data set consisted of 36-step scales, between digital counts of the 0 to 255, for the red, green and blue channels individually as well as in combinations to create a neutral scale. The second data set (labelled: Data Set A) included 793 colours which were a regular sampling of the RGB gamut made up of a $9 \times 9 \times 9$ grid of digital counts from 0 to 255 and a $4 \times 4 \times 4$ grid of digital counts from 5 to 20 in order to increase the sampling in dark colours. The third data set (labelled: Data Set B) had a total of 155 colours from a $4 \times 4 \times 4$ grid of digital counts from 5 to 65, a $4 \times 4 \times 4$ grid of digital counts from 85 to 145 and a $3 \times 3 \times 3$ grid of digital counts from 165 to 235. All of these colours were displayed on the centre area of the display (14 % of full screen) against a mid-grey background having digital counts of 115 for the red, green and blue channels. Their XYZ values were obtained from measurements of the spectral properties using the TSR at a distance of 70 cm from the display. These sets of colours were used either as training data for deriving a model or test data for testing the performance of a model.

4.3.2 Implementation of Characterisation Models

4.3.2.1 GOG Model

As it has described in Section 2.8.3.2, the GOG model includes two stages. The first is a non-linear transform between the digital counts d and the radiometric scalars s for each red, green and blue channel using the model parameters of gain, offset and gamma as this transformation for the red channel is given in Equation 2-45. Similar expressions can be written for the green and blue channels. The second stage is a linear

transformation between the scalars and the XYZ values using a matrix as defined in Equation 2-47.

In order to apply the GOG model, three model parameters have to be derived. To do this, first the primary matrix was obtained by subtracting the XYZ values at zero digital counts from the XYZ values at the maximum output of each red, green and blue channel by setting the XYZ values at zero digital counts as flare in the display. The primary matrix thus determined is given in Equation 4-1. Then, using the inverse of Equation 4-1, the scalar value for each channel for the training data set, either the 33-step neutral scale or the 33-step red, green and blue scales, was estimated. Because of the imperfect system of the display, especially the channel additivity (see Section 4.2.2), the scalars for the peak white ($d_r = d_g = d_b = 255$) do not equal unity. Therefore, the scalars estimated using the primary matrix were normalised by the maximum value of each channel. Three model parameters, gain, offset and gamma, for each channel were then estimated by solving for minimum errors between the estimated scalars for the training data using the inverse of Equation 4-1 and using Equation 2-45. The optimisation was performed using a non-linear optimisation provided by Solver in Microsoft Excel[®]. Consequently, two models were derived; the first based on the 33-step neutral scale; and the second on the 33-steps of the individual red, green and blue scales. Data Set B, consisting of 155 colours, was used to test the performances of these derived models.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 59.98 & 23.34 & 20.69 \\ 31.24 & 64.70 & 13.75 \\ 1.65 & 4.77 & 113.68 \end{bmatrix} \begin{bmatrix} s_r \\ s_g \\ s_b \end{bmatrix} + \begin{bmatrix} 0.46 \\ 0.52 \\ 0.52 \end{bmatrix} \quad \text{Equation 4-1}$$

4.3.2.2 Linear and Polynomial Model

A variety of linear and polynomial regression models with least-squares fitting, similar to those implemented for camera characterisation (see Section 2.8.3.1 and Section 3.4.2.1), were also evaluated. The tested augmented matrices are given in Table 3-9 in Section 3.4.2.1. Although the polynomial models with the transfer matrix up to $M = 3 \times 56$ were evaluated for the camera characterisation (see Section 3.4.2.1), the higher order polynomials were not used (up to $M = 3 \times 20$) in this experiment, since there was acceptable channel additivity and chromaticity constancy of the primaries. All the 33-step scales (red, green, blue and neutral scales) and Data Set A consisting of

the 793 colours were used as training data to derive models. Then, Data Set B (155 colours) was used to test the performance of the derived models.

4.3.2.3 Day, Taplin and Berns Model

A characterisation method proposed by Day, Taplin and Berns was implemented (see Section 2.8.3.3). Similar to the GOG model, it consists of two stages; the first is to characterise each channel's OETF and the second is for the transformation between the radiometric scalars and XYZ values. In this model, the OETFs are described by three one-dimensional look-up tables (LUTs) as expressed in Equation 2-48. A transformation matrix is initially used to describe the relationship between the radiometric scalars and XYZ values as given in Equation 2-47.

In this study, the three one-dimensional LUTs of scalars corresponding to 256 digital counts were created for each red, green and blue channel either from the measurements of the individual red, green and blue 33-step scales or that of the neutral scale using piecewise cubic-spline interpolation. The LUTs were created after the black level had been subtracted from each measurement. Then, the XYZ values of all 33-step scales and Data Set A (793 colours) were predicted using the initial transformation matrix consisting of the maximum output of each channel and the measured XYZ values at zero digital counts as shown in Equation 2-47. Using a non-linear optimisation provided by Solver in Microsoft Excel[®], the coefficients in the transformation matrix were adjusted until the mean colour difference, CIELAB ΔE^*_{ab} , between the measured and estimated XYZ values of all 33-step scales and Data Set A was minimised. Then, the LUTs were re-computed using the adjusted transformation matrix. The model performances were evaluated using the 155 colours of Data Set B.

4.3.3 Display Characterisation Performance

The implemented models and their results are summarised in Table 4-1 for the GOG models, Table 4-2 for the linear and polynomial regression models and Table 4-3 for the Day, Taplin and Berns models.

Table 4-1: Performance of the GOG models.

	Training				Test			
	Training Error (CIELAB ΔE^*_{ab})				Test Error (CIELAB ΔE^*_{ab})			
	Median	Mean	Max	Min	Median	Mean	Max	Min
Data	33-step red, green and blue scales				Data Set B (155 colours)			
	0.40	0.51	1.65	0.051	1.32	1.22	2.32	0.23
Data	33-step neutral scale				Data Set B (155 colours)			
	0.56	0.54	1.19	0.016	0.88	0.94	2.78	0.13

Table 4-2: Performance of the linear and polynomial regression models.

	Training				Test			
Data	All 33-step red, green, blue and neutral scales + Data Set A (793 colours)				Data Set B (155 colours)			
Size of M	Training Error (CIELAB ΔE^*_{ab})				Test Error (CIELAB ΔE^*_{ab})			
	Median	Mean	Max	Min	Median	Mean	Max	Min
3 × 3	20.8	32.9	163.6	0	15.8	20.8	65.0	2.79
3 × 4	23.0	54.2	283.5	1.32	13.8	43.7	232.9	1.47
3 × 5	24.9	54.1	240.5	0.81	13.31	38.4	193.1	1.72
3 × 9	1.85	2.41	11.7	0	1.65	2.61	9.54	0.22
3 × 10	1.92	2.41	11.2	0.058	1.63	2.45	7.98	0.22
3 × 11	1.84	2.38	11.1	0.057	1.64	2.44	7.92	0.22
3 × 20	0.84	1.38	11.5	0.042	1.06	1.71	8.05	0.073

Table 4-3: Performance of the Day, Taplin and Berns models.

	Training				Test			
Data	All 33-step red, green, blue and neutral scales + Data Set A (793 colours)				Data Set B (155 colours)			
LUTs	Training Error (CIELAB ΔE^*_{ab})				Test Error (CIELAB ΔE^*_{ab})			
	Median	Mean	Max	Min	Median	Mean	Max	Min
red, green and blue scales	1.24	1.33	4.53	0	1.08	1.08	2.58	0.14
neutral scale	0.76	0.84	2.81	0	0.71	0.75	2.09	0.17

It can be seen that the GOG models and the Day, Taplin and Berns models generally showed good performance compared with the linear and polynomial models. Although the GOG model is a model for CRT displays, it provided a good fit. This is because the OETFs of the particular LCD had a shape representing a power function. In Figure 4-8, the normalised digital counts in the range 0-1 for each channel are plotted against the scalars from the measurements of the 33-step neutral scale together with that estimated using the parameters from the GOG model in Figure 4-8 (a), and that from the optimised transform matrix and interpolation for the Day, Taplin and Berns model in Figure 4-8 (b). There is only slight dispersion between the measured values (×) and

estimated values from the GOG parameters (full line) at the high digital counts. It can be seen that the display manufacturer has designed this LCD to mimic the typical OETF characteristic of a CRT. But, note that not all LCDs have this type of OETF characteristic; some LCDs have different characteristics such as an S-shape function (Kwak & MacDonald, 2001; Sharma, 2002) and could cause the GOG model to fail.

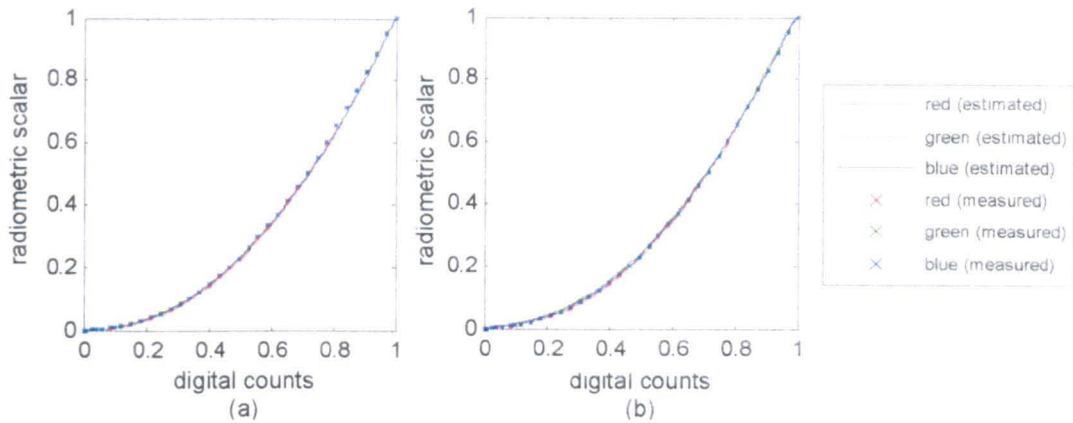


Figure 4-8: The radiometric scalars plotted against the digital counts for the red, green and blue channels. (a) The scalars from the measurements (x) and estimated values (full line) using the parameters from the GOG model. (b) The scalars from the measurements (x) and estimated values (full line) using the optimised transformation matrix from the Day, Taplin and Berns models.

In the GOG and the Day, Taplin and Berns models, the results based on the neutral scale provided better results than those based on the individual red, green and blue channels. If the display had a perfect input and output system, the performance should be the same. However, the display used in this experiment did not have a perfect system (see Section 4.2). The evaluation of the chromaticity constancy revealed the inconstancy of the red and blue channels compared with the green channel. This lack of the constancy was also found from a comparison of the transformation matrices based on the measurements and the optimisation in the Day, Taplin and Berns model using the neutral scale, as shown in Equation 4-2 and Equation 4-4 respectively, and their absolute difference in Equation 4-3. The large changes occurred for the red and blue channels and indicated the lack of chromaticity constancy. It confirms that the optimisation of the matrix is effective and compensates for the lack of the chromaticity constancy.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 59.98 & 23.34 & 20.69 \\ 31.24 & 64.70 & 13.75 \\ 1.65 & 4.77 & 113.68 \end{bmatrix}_{\text{measured}} \begin{bmatrix} s_r \\ s_g \\ s_b \end{bmatrix} + \begin{bmatrix} 0.46 \\ 0.52 \\ 0.52 \end{bmatrix} \quad \text{Equation 4-2}$$

$$|measured - optimised| = \begin{bmatrix} 0.65 & 0.25 & 0.13 \\ 0.30 & 0.03 & 0.24 \\ 0.16 & 0.12 & 0.54 \end{bmatrix} \text{ for } \begin{bmatrix} X_{r,max} & X_{g,max} & X_{b,max} \\ Y_{r,max} & Y_{g,max} & Y_{b,max} \\ Z_{r,max} & Z_{g,max} & Z_{b,max} \end{bmatrix}$$

Equation 4-3

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 59.33 & 23.59 & 20.82 \\ 30.94 & 64.67 & 13.99 \\ 1.91 & 4.89 & 113.14 \end{bmatrix}_{\text{optimised}} \begin{bmatrix} s_r \\ s_g \\ s_b \end{bmatrix} + \begin{bmatrix} 0.46 \\ 0.52 \\ 0.52 \end{bmatrix}$$

Equation 4-4

A possible reason of the better performance of the models from the neutral scale is that these models better compensated the lack of the channel additivity and also cross-talk which affect to colours which are a mixture of the red, green and blue channels, since the neutral colours had already incorporated with these problems. Therefore, the models from the neutral scale showed the better predictions for the test colours the majority of which were mixtures of the red, green and blue channels. In fact, the images of the metallic-coating samples which the model is applied to, also did not have much pure colour. The performance of the GOG and the Day, Taplin and Berns models were very similar; the median test errors were 0.88 and 0.71 in term of CIELAB colour difference ΔE^*_{ab} respectively, but the difference was statistically significant ($p < 0.05$) according to the Wilcoxon signed-rank test (see Section 2.10.5). Consequently, the Day, Taplin and Berns models provided the best performance.

The model was tested only in terms of a “forward model”, which provides XYZ values for given pixel digital counts. In this experiment, the purpose of the characterisation model was to display images whose pixel values were specified in terms of XYZ values. To achieve this, a “inverse model” is required, which provides a mapping from XYZ values to the corresponding RGB digital counts. Although an inverse model is generally expected to work well, if a forward model performs well, it is useful to confirm this. Therefore, the red, green and blue digital counts of the 155 colours in Data Set B (the same set used to test the forward models) were predicted from the measured XYZ values as input using the inverse model of the Day, Taplin and Berns model. Then, these predicted digital counts were displayed on the LCD and measured using the TSR. Finally the colour differences between the input and measured XYZ values were calculated in order to evaluate the inverse model. The median, mean, maximum and minimum values of colour differences were 0.80, 0.89, 2.98 and 0.02 ΔE^*_{ab} respectively. Thus the performance is slightly worse than for the forward model, but still better than the GOG and the linear and polynomial forward

models. Consequently, the Day, Taplin and Berns model derived based on the neutral scale was used in this study.

4.4 Image Reproduction

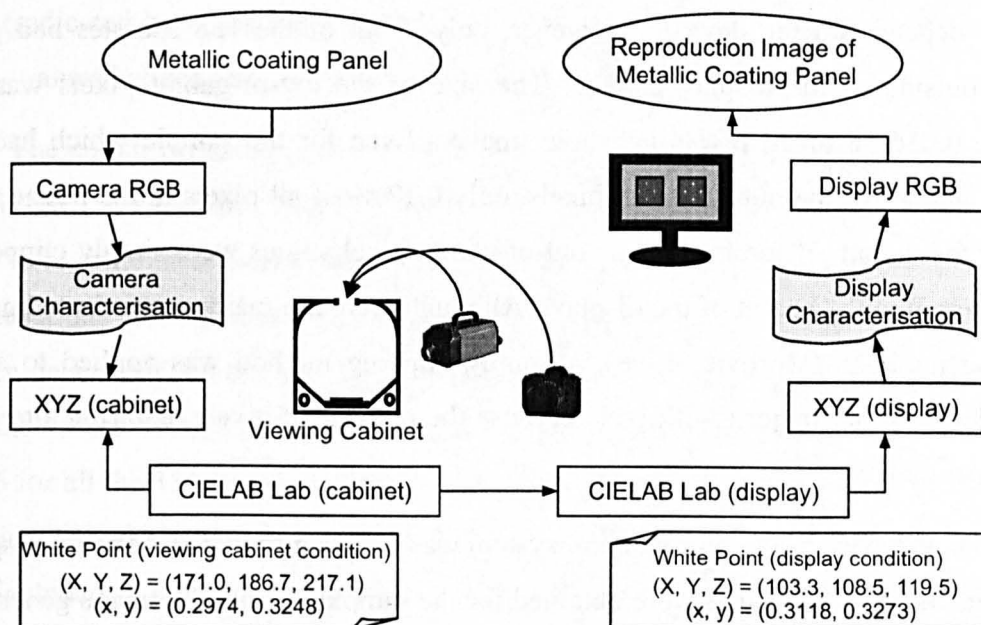


Figure 4-9: A flowchart for image reproduction process.

Images (image samples) were generated from the metallic-coating samples (physical samples) to display on the LCD. The physical samples were the same as the metallic-coating samples used in Chapter 3. A worked flowchart is given in Figure 4-9. The digital images of the physical samples were captured using the digital camera in the viewing cabinet used for the coarseness assessment (see Section 3.4.1) and the camera RGB values of the captured images were transformed to XYZ values corresponding to the viewing cabinet condition where the perceptual coarseness of the physical samples was assessed, via the camera characterisation (see Section 3.4.2.1). Since the light source in the viewing cabinet and the white point of the display were different in their SPD (XYZ values and chromaticity for both conditions are given in Figure 4-9), it was necessary to convert the XYZ values corresponding to the viewing cabinet condition to those corresponding to the display condition. Since D65 was simulated in both conditions and therefore the chromaticity difference was relatively small. Hence, a simple transformation between these two conditions was made in CIELAB colour space.

The CIELAB values, which were transformed from XYZ values corresponding to the white point of the viewing cabinet condition, were converted back to XYZ values, but for the display condition according to the white point of the LCD. Then, the RGB values for the display were obtained through the display characterisation model (see Section 4.3). During the transformation process, it was found that some colours were outside of the gamut of the display, since all the devices have a limited gamut and its volume depends on the device. However, only 7 out of the 156 samples had pixel values outside of the display gamut. The size of the out-of-gamut pixels was, on average 0.036 % of all pixels in whole image. Even for the sample which had the largest number of the out-of-gamut pixels, only 0.12 % of all pixels in the image were outside the gamut. Therefore, those out-of-gamut pixel values were simply clipped to bring them into the gamut of the display. Although there are many sophisticated gamut mapping methods (Morovič, 1998), a simply clipping method was applied to avoid changing the other majority of pixels because the number of pixels requiring alteration was very small.

The accuracy of the reproduction system was evaluated in CIELAB colour space. Four sets of CIELAB values were obtained for the samples during the image generation process (Figure 4-9). The first set was the values representing the physical samples and they are labelled MP (Measured Physical samples). These were obtained based on the measurements of the physical samples in the viewing cabinet using the TSR (see Section 3.4.1). The second set of CIELAB values was based on the captured images of those samples. The camera RGB values were transformed to XYZ values and then to CIELAB values. As a result, the CIELAB values for each pixel in the images were obtained. The mean CIELAB values of the pixels in each image were taken as representative of each sample and they are called EC (Estimated by Camera characterisation). The third data set, EL (Estimated for LCD), was based on the input XYZ values to the display; namely they were XYZ values transformed from EC according to the white points of the cabinet and display conditions. Similar to EC, the mean values of the pixels in each image represented the each sample. The fourth data set, MI (Measured Image samples), comprised the CIELAB values based on the displayed image samples; the generated image samples were displayed on the screen and measured using the TSR from a distance of 70 cm from the display. The differences in these data sets are that MP and MI were based on the measurements using the TSR, while the EC and EL were based on the computational transformations. The

colour differences, CIELAB ΔE^*_{ab} , between MP, EC, EL and MI are summarised in Figure 4-10. Figure 4-10 (a) shows the ΔE^*_{ab} differences between MP and the other sets (EC, EL and MI). Each bar represents a range between maximum and minimum ΔE^*_{ab} value for all the 156 samples (red bar), 5 grey colour samples (grey bar), 50 purple colour samples (purple bar), 50 green colour samples (green bar) and 50 blue colour samples (blue bar). The graph also shows the median values associated with the data (indicated by a dot on each bar). Similarly, Figure 4-10 (b, c, and d) show the ΔE^*_{ab} values between EC, EL or MI and the other data sets.

The colour difference between the MP and EC was a median ΔE^*_{ab} value of 0.38 for all the 156 samples. This can be considered reasonable difference, because the EC data were estimated using the camera characterisation model which includes some optimisations (see Section 3.4.2.1), since there are often no exact solutions to correlate device-dependent and device-independent values. It can be seen clearly from the graphs that the values of ΔE^*_{ab} between EC and EL are negligible; a median ΔE^*_{ab} value was 0.036 for all the 156 samples. It is resulted of the mathematical transformation in order to adjust the white point from the cabinet to the display conditions, so no any estimation was included in this transformation. Consequently, the difference between MP and EL (a median ΔE^*_{ab} value of 0.39) was similar to that of between MP and EC. However, it can be seen that the differences between MI with the other sets are relatively large; median ΔE^*_{ab} values for all 156 samples were 4.65 between MP and MI, 4.63 between EC and MI and 4.64 between EL and MI. These large errors were reasonable if significant errors were generated in the display characterisation process. However, as has been mentioned in Section 4.3, the display characterisation errors were reasonably small. The median value of ΔE^*_{ab} was 0.8 with the inverse model according to the evaluation using a set of the 155 test colours. These test colours were widely distributed in RGB colour space. Therefore, it can be assumed that the model was able to satisfactorily predict the coordinates of colours over the required range. Further examination was carried out by comparing CIELAB L^* , C^* and a^*b^* values of the MP and MI data sets shown in Figure 4-11 (a, b and c) respectively. These graphs reveal that the colour differences were caused not so much by the lightness shifts, but by the chroma and hue shifts. In fact, the grey samples had smaller errors as has already been demonstrated in Figure 4-10. A possible reason of the errors could be traced by examining the characteristic difference between the test colours (used to test the display characterisation model) and the samples. The test colours were uniform patches, while

the samples were not uniform. In order to demonstrate the influence of the spatial configuration, a uniform grey patch, a checked image and a striped images consisting of alternating rows or columns of black and white with one pixel each cycle were displayed on the LCD (Rhodes, 2007). The checked and striped images are illustrated in Figure 4-12. It was found that, although these images had a same portion of black and white, they looked different, for example, one was more reddish or greenish than the others. This cannot be explained solely by the lack of the spatial independency which showed colour shifts toward the surrounding colours (see Section 4.2.4). This defect indicates that colours cannot be reproduced accurately on a pixel-by-pixel basis. Moreover, the degree and direction of the errors are different depending on the spatial arrangement, and also depending on displays (It was found by demonstrating this with several displays). It seems that this influence is larger if the colour differences between neighbouring pixels are large. This can be attributed to the fact that the accuracy of the green colour samples was worst followed by the blue then the purple colour samples in terms of ΔE^*_{ab} (Figure 4-10). According to the visual assessment results using the physical samples, it was found that the perceptual coarseness of the average green colour samples was larger than that of the blue and purple colour samples (Figure 3-18). There should have been larger variation in the pixel values in coarse samples, *i.e.*, in the green colour samples. Consequently, it can be presumed that one reason for the large colour differences between MP and MI (Figure 4-10) is originated in the defect of physical structure of the displays. Namely, it is a limitation of LCD technology. Although the accuracy of image generation was not excellent, the problems were accepted for this study.

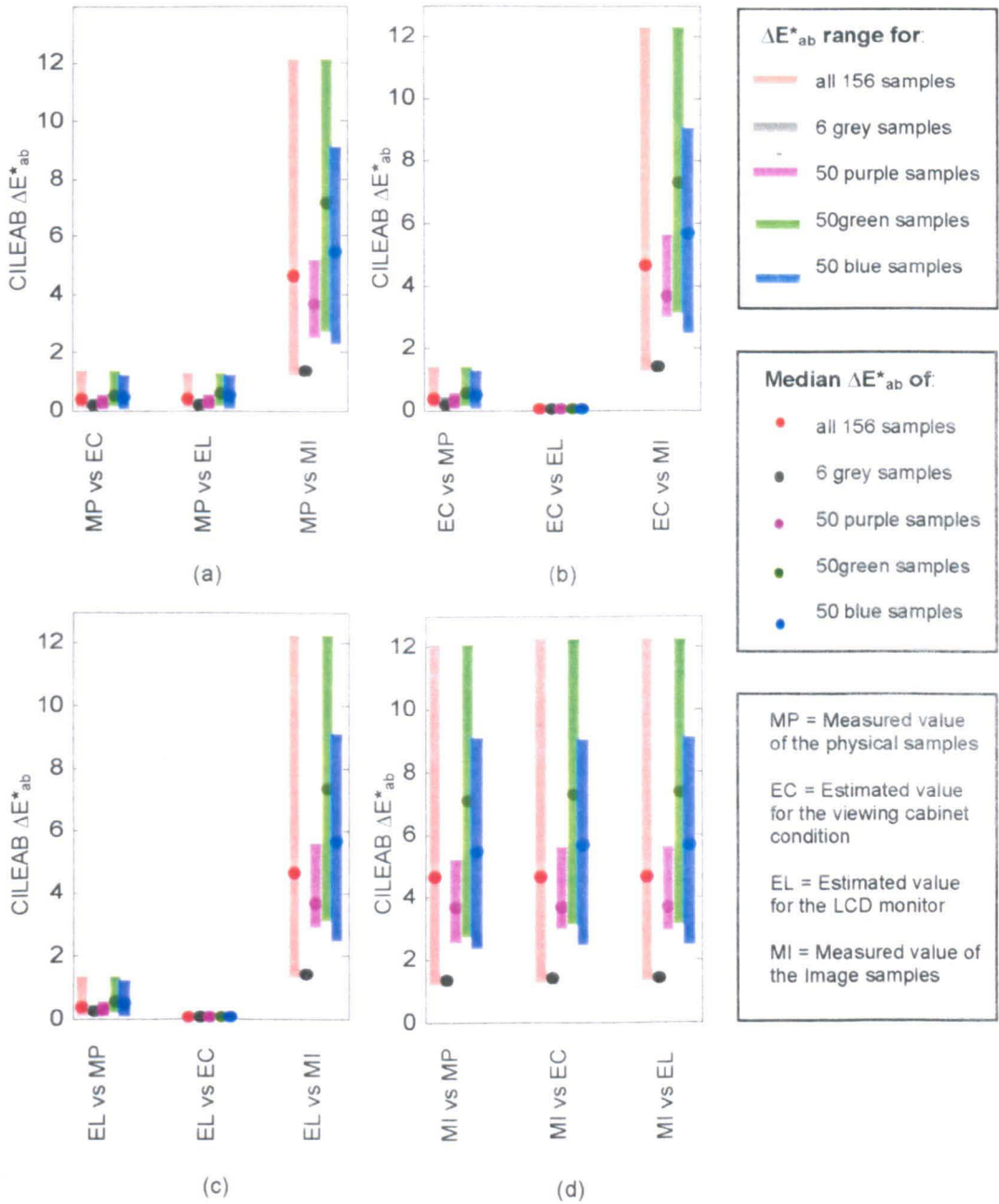


Figure 4-10: CIELAB colour difference between data sets MP, EC, EL and MI. (a) Comparisons of MP with EC, EL and MI. (b) Comparisons of EC with MP, EL and MI. (c) Comparisons of EL with MP, EC and MI. (d) Comparisons of MI with MP, EC and EL.

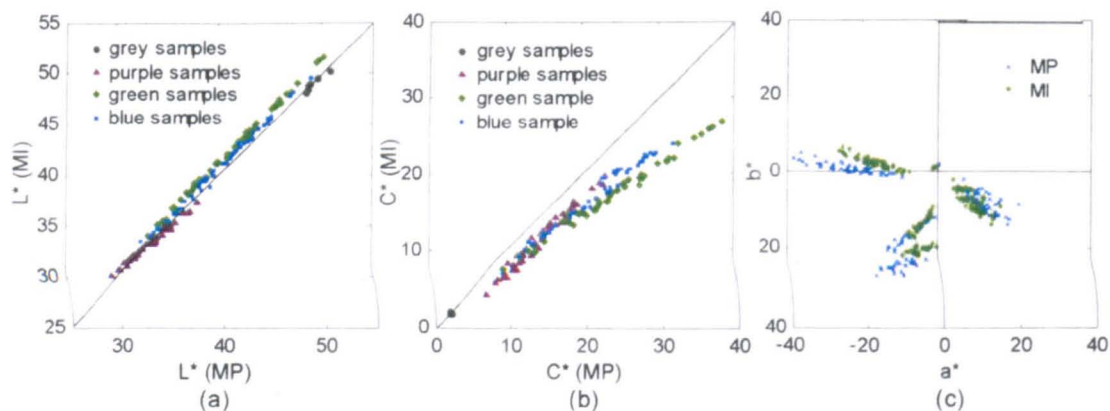


Figure 4-11: CIELAB (a) L^* (b) C^* and (c) a^*b^* differences between MP and MI.

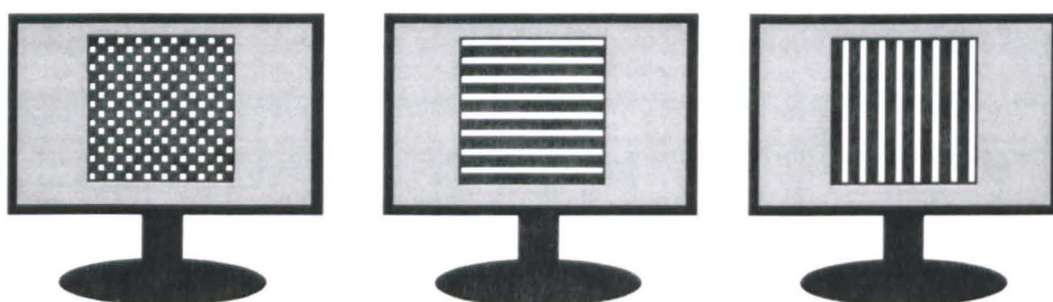


Figure 4-12: Images used to test the errors of the output colours. The colour of the three images looks different, although they have the same ratio of black to white pixels.

4.5 Visual Assessment using Image Samples

Although the resolution of the camera was not high enough to resolve individual aluminium flakes (see Section 4.1), the resolution of the display is even lower. In order to display the image sample with the same size as the physical sample (8×8 cm), the image had to be captured with a lower resolution which gave a $249 \times 249 \mu\text{m}$ area in each pixel, since the physical size of one pixel of this display was $249 \times 249 \mu\text{m}$. In this case, most of the detail has been lost, because information read out from a pixel of the camera was integrated over a larger area. This resulted in loss detail in the perceptual texture as seen in the images. Therefore, the images were captured at a higher resolution to obtain much more details. This resulted in an image size of 21×21 cm on the display and a viewing distance of 140 cm, in order to keep the same angular display size of $8.5^\circ \times 8.5^\circ$ for the experimental conditions for both the physical samples and the image samples. The specification of this viewing condition is given in

Table 4-4 as Condition 1. Note that the conditions for the coarseness assessment of the physical samples were a sample size of 8×8 cm, a viewing distance of 54 cm and an angular display size of $8.5^\circ \times 8.5^\circ$ (see Section 3.3.2.1). It was found that the appearance of the images in Condition 1, however, did not perfectly match that of the physical samples. This is mainly because of the errors that occurred during the image generation process, as has been discussed in the previous section, and also the loss of detail in the imaging-forming system (both of the camera and the display) as characterised by the modulation transfer factor. Moreover, the study by Johnson and Montag (2005) showed a disagreement in perceptual appearance of images having the same retinal size, if the viewing distances are different. The appearance of coarseness in the images in Condition 1 was such that was smaller than that of the physical samples because of the increased viewing distance. In common practice, it is usual to move closer to an object to see its detail. Therefore, to achieve image samples having the perceptual coarseness equivalent to that of the physical samples, the coarseness of the image samples and the physical samples were compared in various conditions (e.g. various viewing distances). Since it was difficult to simply compare the “absolute” coarseness of the image samples with the physical samples because not only their coarseness but also their colours were different, the “relative” coarseness was investigated, *i.e.*, the coarseness difference of two physical samples placed in a viewing cabinet was compared with that of two image samples on the display at nine viewing distances as shown in Table 4-4, in order to find conditions in which coarseness differences appeared similar. Then, according to the results of this coarseness difference assessment in varying conditions, the coarseness of the image samples was scaled by observers under the condition which showed the closest perceptual coarseness differences between the image and physical samples. Additionally, the coarseness of the image samples was scaled using Condition 1 which was relative reproduction of the conditions used in the visual assessment of the physical samples. The following sections describe these three visual assessments.

Table 4-4: The viewing conditions for the visual assessments.

Condition	Image size (pixel)	Image size (cm)	Viewing distance (cm)	Angle display size (degree)	Magnification ratio (%)
1	832 x 832	20.7	140	8.5 x 8.5	100
2	714 x 714	17.8	120	8.5 x 8.5	117
3	595 x 595	14.8	100	8.5 x 8.5	140
4	476 x 476	11.9	80	8.5 x 8.5	175
5	417 x 417	10.4	70	8.5 x 8.5	200
6	357 x 357	8.9	60	8.5 x 8.5	233
7	298 x 298	7.4	50	8.5 x 8.5	280
8	238 x 238	5.9	40	8.5 x 8.5	350
9	179 x 179	4.4	30	8.5 x 8.5	466

4.5.1 Comparison of Coarseness Differences

Ten observers, six females and four males, aged between 25 and 35, compared the coarseness difference between two physical samples with the coarseness difference between two image samples. The two physical samples were the coarsest grey sample and the second least coarse grey sample according to the results of the visual assessments using the physical sample described in Section 3.3. The two image samples were the reproduced images of those two selected physical samples. The physical samples were presented in a viewing cabinet beside the display on which the image samples were displayed, as illustrated in Figure 4-13 for Condition 1 and Condition 6 as examples. The assessments were carried out for the nine conditions listed in Table 4-4; namely, the image samples were observed at the nine different viewing distances. Since the angular display size of the image was kept constant for all the conditions, the actual image size (physical size) was made smaller by showing only part of the full image as illustrated in Figure 4-13. Accordingly, the images appeared magnified to the observers as the viewing distance decreased, although the image itself did not change. The cabinet where the physical samples were placed in this assessment, was different from the one used for the coarseness assessments for the physical samples in Section 3.3. In order to produce the luminance level to be the same as that of the display (100 cd/m^2), a cabinet whose luminance level was adjustable was used. Other viewing conditions were such that the sample size was $8 \times 8 \text{ cm}$ and the viewing distance was 54 cm which was consistent with the coarseness assessment conditions used in Section 3.3.

The magnitude estimation method (see Section 2.9.2) was applied for scaling. Observers were asked to assign a number that best described the coarseness difference

between the two image samples when compared with the coarseness difference between two physical samples whose coarseness difference was assigned a value of five. For example, if an observer scaled the coarseness difference of a pair of image samples as five, it meant that the coarseness difference of these two image samples was the same as that of the two physical samples. The geometric mean of raw observer data was calculated as a measure of the central tendency of their data for each condition and also a general approximation of the standard deviation of scale values was computed using Equation 4-5 (Engeldrum, 2000).

$$\sigma \approx \left[\frac{\bar{O}_i^2}{n-1} \sum_{j=1}^n [\ln(O_{ij}) - \ln(\bar{O}_i)]^2 \right]^{0.5} \quad \text{Equation 4-5}$$

where O_{ij} is the raw data of the observer j ($j=1,2,\dots,n$) for the observation condition i , \bar{O}_i is the geometric mean of the observers data. The geometric means are presented in Figure 4-14 together with ± 1 standard deviation from the means (error bars). A scale value of 4.8 obtained in Condition 6 was the closest match to the coarseness difference of the physical samples which was assigned a value of five, but the error bars indicates the insignificance of the difference between the coarseness difference of the two physical samples and that of the two image sample in any condition. However, the standard deviation calculated using Equation 4-5 assumes the log normal distribution of the observer data. The observer data obtained from this experiment do not entirely satisfy this assumption. Therefore, the Wilcoxon signed-rank test was performed and the significances of the differences (p-values) between the coarseness difference of the two physical samples and those of two image samples in each Condition 1 to Condition 9 are given in Table 4-5. This table shows that, although there are no significant differences in most conditions, the coarseness difference was significantly smaller in Condition 1 ($p<0.05$) which was the relative reproduction of the viewing conditions to the coarseness assessment described in Section 3.3. Consequently, since Condition 6 was the closest match with the physical samples, this was selected as an optimal condition to scale the perceptual coarseness of the image samples.

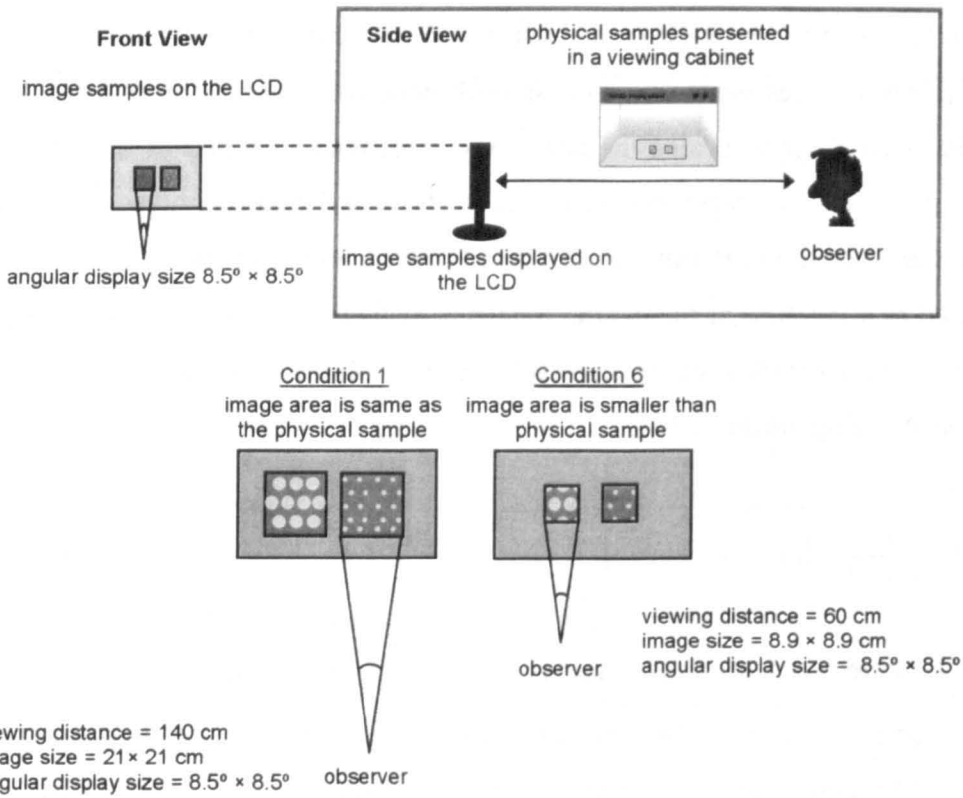


Figure 4-13: The viewing geometry used for comparison of the physical samples with the image samples (top) and the examples of the image sample on the LCD and parameters for Condition 1 and 6 (bottom).

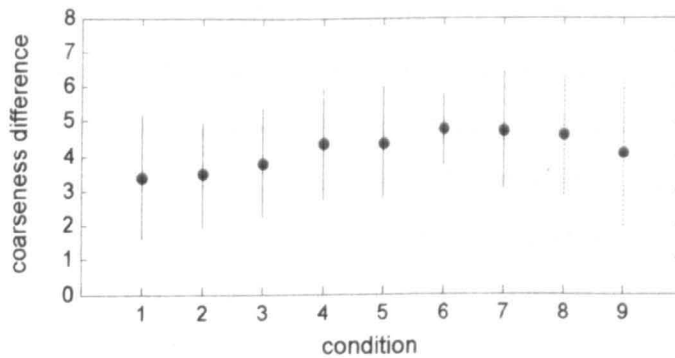


Figure 4-14: Geometric mean perceptual coarseness difference for each of nine conditions together with the error bars. Coarseness difference five indicates that a pair of image samples displayed on the display had an equivalent coarseness difference to the physical sample pair presented in the viewing cabinet.

Table 4-5: P-values by the Wilcoxon signed-rank test to investigate the significance of the differences between the coarseness difference of two physical samples and those of two image samples in each Condition 1, 2, 3, 4, 5, 6, 7, 8 and 9.

Condition	1	2	3	4	5	6	7	8	9
p-value (1-tailed)	0.049	0.033	0.101	0.179	0.312	0.367	0.477	0.459	0.286

4.5.2 Scaling Perceptual Coarseness in Condition 6

The experimental procedure was similar to the visual assessment of the physical samples carried out in Section 3.3.3. Two image samples were displayed on the LCD for each trial. One was a reference image and the other a test image. The reference image was the image of the physical sample used as a reference sample in the coarseness assessment in Section 3.3.3. The 1-9 categorical judgement scaling method with a reference was applied (see Table 3-4). The difference from the previous experiment, Section 3.3.3, was that the observers were allowed to give the coarseness values with up to two decimal places, *e.g.*, 5.25, 5.5 or 6. All 156 test images were presented in a random order. To check the repeatability, each observer carried out the assessment twice. Fourteen observers, including six females and eight males aged between 25 and 35, participated in the visual assessment. A total of 4368 (14 observers × 2 sessions × 156 samples) categorical judgments were made.

4.5.3 Scaling Perceptual Coarseness in Condition 1

The coarseness assessment was also conducted using Condition 1. Condition 1 was the relative condition to that used for assessing the physical samples (described in Section 3.3.3). Thus, the area of the image samples displayed was the same as that of the physical samples presented to the observers. The same categorical judgement scaling method, as used for Condition 6, was applied. In this experiment, 66 image samples out of the 156 image samples were used. All 6 grey samples and 20 from each of the blue, green and purple colour samples were selected. Ten observers, six females and four males aged between 25 and 35, participated in this experiment. A total of 1320 by 10 observers (2 sessions for each observer to check their repeatability for 66 samples) categorical judgments were made.

4.6 Data Analysis and Result

The mean-category value method (see Section 2.9.1) was employed to represent the observer data for each sample. Therefore, arithmetic means of the raw observer data were calculated as scale values of each sample for both the Condition 1 and Condition 6 experiments and they are given in Appendix II. In the following sections, observer variability (observer repeatability and accuracy) were investigated for each condition.

Then, the quality of the image samples was assessed by analysing the agreement and correlation between the scale values obtained from the three conditions: the visual assessments using the image samples in Condition 1 and Condition 6, and also the physical samples in the viewing cabinet (Section 3.3). In addition, the scale values were also compared to the predicted coarseness values from the image samples for each experimental condition using the computational model developed in Section 3.4.

4.6.1 Observer Variability

Observer variability was investigated for two aspects: observer repeatability and observer accuracy. The investigations of these were carried out in terms of R^2 and CV. Summaries of the observer variability; the mean, median, maximum and minimum for Condition 6, Condition 1, and also for the coarseness assessment in the viewing cabinet using the physical sample (labelled: Physical Sample), are given in Table 4-6 for the repeatability and Table 4-7 for the accuracy. In addition, each observer's repeatability is shown in Table 4-8 and Table 4-9 for Condition 6 and Condition 1 respectively. Similarly, the accuracy obtained in each session for Condition 6 and Condition 1 can be found in Table 4-10 and Table 4-11 respectively.

It can be seen from the results of observer repeatability and accuracy for Condition 6 that the observer variability for Condition 6 shows a similar tendency to the results for the physical samples. The both repeatability and accuracy was very good for the grey samples in terms of R^2 , but not in terms of CV. This indicates that there was a linear correlation, however the observers did not give the same categories to the samples. Also, the result for the purple colour samples showed the poorest performance; although their correlations in terms of R^2 are similar to those of the blue samples, the CV values indicate poorer performance than for the blue samples. A possible reason of the better performance for the grey colour samples might possibly be that there are only five grey samples with each relatively distinguishable coarseness difference. On the other hand, the poor performance for the purple colour samples could be attributable to the relatively small distribution of coarseness and to the low coarseness levels (Figure 3-18), and therefore it might be difficult for the observers to distinguish between the samples because of the limitations in the inability of our spatial vision system (see Section 2.2.3).

The observer repeatability and accuracy of individual blue, green and purple colour samples obtained in Condition 1 generally showed better performance than that

in Condition 6 and the physical samples (Table 4-6). Usually, the observer variability indicates the reliability of the assessments. However, it is difficult to say whether Condition 1 is superior to the other conditions, since the overall variability for all the samples were equivalent to the other conditions and the number of the samples used in Condition 1 was less than that used in the other conditions so there was less opportunity for confusion between the coarseness levels of the samples.

Consequently, it can be seen that the repeatability and accuracy for all the samples are similar for all conditions. This indicates that observers assessed the coarseness with equivalent accuracy under these three conditions, although the absolute appearances of the coarseness at these conditions were different, *i.e.*, there was a much smaller perceived coarseness difference under Condition 1 than that under Condition 6, or using the viewing cabinet, according to the experimental results presented in Figure 4-14.

Table 4-6: Observer repeatability: mean, median, maximum and minimum values of all observer data for coarseness assessment using Condition 6 and Condition 1, and the physical samples (from Chapter 3: Table 3-6).

Samples	R ²					CV				
	All	Grey	Blue	Green	Purple	All	Grey	Blue	Green	Purple
Condition 6										
Mean	0.69	0.80	0.39	0.46	0.34	14	14	13	10	21
Median	0.73	0.87	0.40	0.45	0.37	15	12	14	9	20
Max	0.83	0.98	0.70	0.69	0.59	19	29	22	14	37
Min	0.34	0.38	0.10	0.20	0.04	8	5	8	7	9
Condition 1										
Mean	0.68	0.68	0.50	0.51	0.37	12	15	11	9	15
Median	0.72	0.79	0.51	0.53	0.32	13	13	10	9	14
Max	0.84	0.93	0.61	0.76	0.67	15	35	17	12	23
Min	0.31	0.10	0.31	0.15	0.11	6	7	6	5	8
Physical Sample										
Mean	0.69	0.90	0.32	0.37	0.35	18	14	17	13	28
Median	0.72	0.92	0.36	0.34	0.36	17	12	18	13	26
Max	0.79	0.98	0.43	0.57	0.57	24	22	20	15	54
Min	0.53	0.74	0.17	0.26	0.14	12	8	10	8	14

Table 4-7: Observer accuracy: mean, median, maximum and minimum values of all observer data for the coarseness assessment using Condition 6 and Condition 1, and the physical samples (from Chapter 3: Table 3-7).

Samples	R ²					CV				
	All	Grey	Blue	Green	Purple	All	Grey	Blue	Green	Purple
Condition 6										
Mean	0.81	0.89	0.58	0.66	0.56	12	12	11	9	18
Median	0.84	0.92	0.58	0.67	0.60	12	12	11	9	19
Max	0.90	0.99	0.76	0.82	0.73	18	26	17	15	39
Min	0.49	0.55	0.25	0.39	0.07	7	5	6	4	9
Condition 1										
Mean	0.79	0.79	0.67	0.67	0.57	11	12	10	10	14
Median	0.83	0.89	0.68	0.70	0.56	12	10	10	10	13
Max	0.89	0.97	0.89	0.87	0.87	16	42	17	16	28
Min	0.36	0.21	0.48	0.25	0.26	7	5	5	5	6
Physical Sample										
Mean	0.82	0.93	0.55	0.61	0.58	14	12	13	10	21
Median	0.83	0.96	0.57	0.63	0.61	14	10	13	10	20
Max	0.88	0.99	0.76	0.81	0.78	18	22	19	14	33
Min	0.68	0.58	0.31	0.32	0.23	10	5	8	7	14

Table 4-8: Observer repeatability for each observer using Condition 6.

Observer	R ²					CV				
	All	Grey	Blue	Green	Purple	All	Grey	Blue	Green	Purple
1	0.71	0.58	0.46	0.66	0.36	16	26	14	8	27
2	0.67	0.97	0.11	0.37	0.40	19	9	22	9	30
3	0.48	0.59	0.24	0.29	0.25	16	13	20	14	14
4	0.79	0.98	0.52	0.45	0.59	8	5	8	8	9
5	0.83	0.94	0.55	0.57	0.38	10	6	9	8	16
6	0.69	0.74	0.51	0.45	0.23	12	11	9	10	18
7	0.34	0.56	0.11	0.22	0.04	15	25	15	9	21
8	0.70	0.80	0.25	0.20	0.47	15	18	14	12	21
9	0.76	0.98	0.70	0.61	0.47	16	9	13	11	27
10	0.81	0.89	0.42	0.69	0.41	9	12	8	7	10
11	0.76	0.98	0.56	0.61	0.18	15	10	15	7	30
12	0.77	0.85	0.35	0.57	0.39	14	16	15	9	19
13	0.75	0.94	0.29	0.41	0.28	17	9	13	11	37
14	0.64	0.38	0.38	0.31	0.31	15	29	13	13	18

Table 4-9: Observer repeatability for each observer using Condition 1.

Observer	R ²					CV				
	All	Grey	Blue	Green	Purple	All	Grey	Blue	Green	Purple
1	0.72	0.58	0.59	0.65	0.67	14	22	14	9	16
2	0.64	0.10	0.31	0.53	0.56	15	21	17	10	18
3	0.31	0.38	0.46	0.15	0.11	14	35	10	12	12
4	0.80	0.81	0.47	0.60	0.58	6	7	6	5	8
5	0.84	0.89	0.59	0.76	0.35	12	11	9	6	23
6	0.72	0.83	0.57	0.37	0.19	9	9	9	8	10
7	0.69	0.69	0.48	0.53	0.28	9	14	8	7	12
8	0.79	0.84	0.54	0.44	0.14	14	10	15	9	21
9	0.57	0.93	0.34	0.44	0.17	13	11	14	12	15
10	0.74	0.76	0.61	0.59	0.61	10	14	9	8	10

Table 4-10: Observer accuracy for each session using Condition 6.

Observer & Session		R ²					CV				
		All	Grey	Blue	Green	Purple	All	Grey	Blue	Green	Purple
1	1	0.84	0.93	0.72	0.82	0.66	14	13	13	10	21
	2	0.87	0.72	0.76	0.76	0.66	14	21	12	11	19
2	1	0.83	0.95	0.56	0.56	0.46	18	12	17	15	25
	2	0.80	0.97	0.42	0.70	0.56	16	6	16	14	23
3	1	0.55	0.74	0.43	0.51	0.33	16	16	17	12	19
	2	0.59	0.89	0.48	0.61	0.24	14	10	13	8	22
4	1	0.84	0.96	0.65	0.57	0.65	12	11	12	13	10
	2	0.89	0.91	0.69	0.75	0.73	12	15	11	14	9
5	1	0.90	0.95	0.67	0.81	0.59	9	6	8	8	12
	2	0.90	0.93	0.71	0.68	0.73	7	7	6	7	9
6	1	0.84	0.90	0.69	0.63	0.53	9	10	7	8	12
	2	0.81	0.94	0.58	0.79	0.38	10	10	8	8	15
7	1	0.61	0.89	0.25	0.54	0.45	14	11	13	10	19
	2	0.49	0.79	0.28	0.39	0.07	16	14	10	8	27
8	1	0.88	0.92	0.58	0.74	0.71	10	12	11	6	15
	2	0.84	0.86	0.61	0.46	0.71	12	13	10	9	20
9	1	0.80	0.99	0.69	0.69	0.70	14	16	13	9	23
	2	0.85	0.97	0.64	0.79	0.67	12	12	11	8	21
10	1	0.85	0.91	0.51	0.74	0.55	10	9	9	10	11
	2	0.87	0.86	0.57	0.67	0.67	8	10	8	7	10
11	1	0.81	0.91	0.56	0.68	0.36	15	20	13	6	30
	2	0.89	0.89	0.70	0.80	0.57	10	15	7	4	22
12	1	0.85	0.94	0.58	0.62	0.58	12	14	12	8	18
	2	0.85	0.97	0.62	0.63	0.62	11	5	11	9	14
13	1	0.85	0.94	0.54	0.67	0.60	16	8	10	8	39
	2	0.84	0.91	0.51	0.66	0.54	12	9	9	7	24
14	1	0.82	0.55	0.68	0.65	0.64	11	26	9	9	12
	2	0.82	0.93	0.55	0.51	0.61	12	7	12	11	13

Table 4-11: Observer accuracy for each session using Condition 1.

Observer & Session	R ²					CV					
	All	Grey	Blue	Green	Purple	All	Grey	Blue	Green	Purple	
1	1	0.83	0.72	0.74	0.87	0.81	12	10	14	9	16
	2	0.84	0.97	0.75	0.75	0.87	11	17	8	8	16
2	1	0.83	0.77	0.67	0.53	0.87	16	14	16	16	17
	2	0.75	0.33	0.56	0.66	0.71	16	20	17	13	17
3	1	0.36	0.21	0.71	0.25	0.26	16	42	12	12	14
	2	0.56	0.38	0.58	0.44	0.45	13	16	14	13	12
4	1	0.86	0.94	0.52	0.67	0.72	10	10	10	11	6
	2	0.86	0.87	0.66	0.73	0.76	10	8	10	12	7
5	1	0.89	0.94	0.73	0.86	0.51	12	10	10	10	17
	2	0.87	0.95	0.68	0.71	0.59	15	8	11	11	28
6	1	0.81	0.91	0.74	0.52	0.49	8	10	6	7	10
	2	0.87	0.91	0.8	0.69	0.58	7	7	6	5	9
7	1	0.78	0.89	0.54	0.76	0.32	8	8	8	7	11
	2	0.80	0.81	0.68	0.65	0.54	8	12	7	7	9
8	1	0.87	0.74	0.67	0.84	0.49	13	14	12	5	25
	2	0.83	0.95	0.62	0.55	0.33	12	10	8	8	22
9	1	0.68	0.93	0.48	0.59	0.46	11	5	11	10	13
	2	0.84	0.88	0.89	0.70	0.40	8	10	5	7	10
10	1	0.79	0.84	0.75	0.76	0.58	12	12	12	13	9
	2	0.85	0.92	0.72	0.83	0.59	9	5	8	9	10

4.6.2 Comparisons of Perceptual Coarseness Scaled in Different Conditions

Investigations were carried out on the perceptual coarseness scaled at three conditions: Condition 6 and Condition 1 using the image samples, and the viewing cabinet condition using the physical samples. Figure 4-15 shows the comparisons of the scale values obtained in these conditions together with the measures of their agreement computed in terms of R² and CV. Black lines in each graph are an indication of perfect agreement between the data sets.

The results for all the samples from Condition 6 and Condition 1 are compared in Figure 4-15 (a). The results of an R² value of 0.95 and of a CV value of 5 found between Condition 6 and Condition 1 indicate the excellent agreement. The statistical analysis also shows the insignificant difference (p>0.05) between them. The comparisons of individual colour samples are given in Figure 4-16. While the grey, blue and green colour samples showed excellent agreement (R² values of 0.95, 0.93 and 0.95 and CV values of 11, 3 and 5 respectively), the purple colour samples had a

slightly poor agreement (an R^2 value equal to 0.75 and a CV value of 9). The reason of the poor agreement for the purple colour samples might be originated in the observer variability. The observer repeatability and accuracy for the purple colour samples were poorer than that especially for blue and green colour samples in both conditions. However, overall results indicate, as well as the observer variability (in Section 4.6.1), that the observers could fairly assess the coarseness in these conditions, although absolute coarseness for the far distance (Condition 1) was smaller than that for the closer distance (Condition 6). Under Condition 6, only a part of the sample was presented to the observers comparing with the physical sample and the image sample in Condition 1; Condition 6 had less than half the area compared with the other conditions (see Figure 4-13 and Table 4-4). The results suggest that for assessing the equally distributed fine detail over a sample like a metallic-coating panel used in the experiment, the area presented to the observers are not so important. Observers tend to focus on only a part of the sample and not on the whole sample.

The comparison between Condition 6 and the physical samples is shown in Figure 4-15 (b) and that between Condition 1 and the physical samples was Figure 4-15 (c). The statistical analysis indicates the significant difference between Condition 6 and the physical samples ($p < 0.05$). Although there is no significant difference between Condition 1 and the physical sample for a 5 % significance level, the low p-value ($p = 0.065$) is found. This implies the disagreement between them. The detail can be examined from Figure 4-17 and Figure 4-18. Figure 4-17 compares the results of individual grey, blue, green and purple colour samples of Condition 6 against those of the physical samples. Similarly, Figure 4-18 is for between Condition 1 and the physical samples. In both cases, it can be clearly seen that there is excellent agreement for the higher coarseness scale values, *i.e.*, the results for the blue and green colour samples, but slightly poorer agreement is found for the samples with smaller coarseness scale values, *i.e.*, the results for the purple colour samples. In terms of the image reproduction errors summarised in Figure 4-10 and Figure 4-11, the purple colour image samples are more accurately reproduced than the blue and green colour image samples. Therefore, higher agreement would be expected. But, it must be noted the poor observer repeatability and accuracy for the purple colour samples than that for the blue and green colour samples in Condition 6, Condition 1 and importantly also the visual assessments using the physical samples. Therefore, the scale values for the purple colour samples are less reliable than that for the others. Accordingly, the high

correlation of the scale values between the image and physical samples for the blue and green colour samples verified the quality of the reproduced image samples in respect to their perceptual coarseness, but the poor correlation of the scale values of the purple colour samples cannot be a strong evidence of a failure in the image reproduction.

Overall, the significant difference was found between Condition 6 and the physical samples and slightly poor agreement between the conditions for the samples whose coarseness levels were relatively small. However, for visual assessments in general, the correlations obtained between Condition 6 and the physical samples (an R^2 value of 0.92 and a CV value of 9) and that between Condition 1 and the physical samples (an R^2 value of 0.89 and a CV value of 11) for all the samples can be considered as excellent agreement.

It should be noted that a significant colour shift was found in the generated image samples. The lightness of the images was relatively preserved compared with the errors occurred in the chroma and hue (Figure 4-11). The agreement of the scale values in three conditions suggests the importance of the lightness for the perceptual coarseness rather than the chroma and hue, and also supports the coarseness model to use only the luminance channel to predict perceptual coarseness (see Section 3.4).

Importantly, the agreement of the scale values in three conditions also indicates that the resolution of the images used in this experiment was high enough, even though each pixel size of the camera setting was larger than the size of the aluminium flakes. As was expected, although the coarseness of the metallic-coating samples was originated in the flakes, the observers did not view the flakes themselves; rather they perceived coarseness as a result of the interaction of reflected light and the composition of the coatings, and this caused the coarseness effect to exceed the physical size of the flakes.

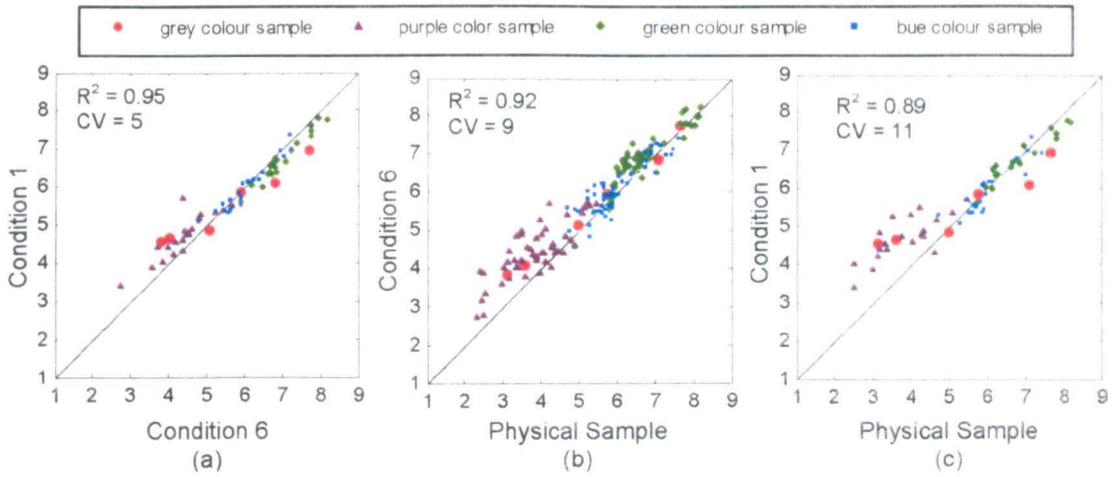


Figure 4-15: Comparisons between the scale values in Condition 6 using the image samples (labelled Condition 6), in Condition 1 (labelled Condition 1) and in the viewing cabinet using the physical samples (labelled Physical Sample).

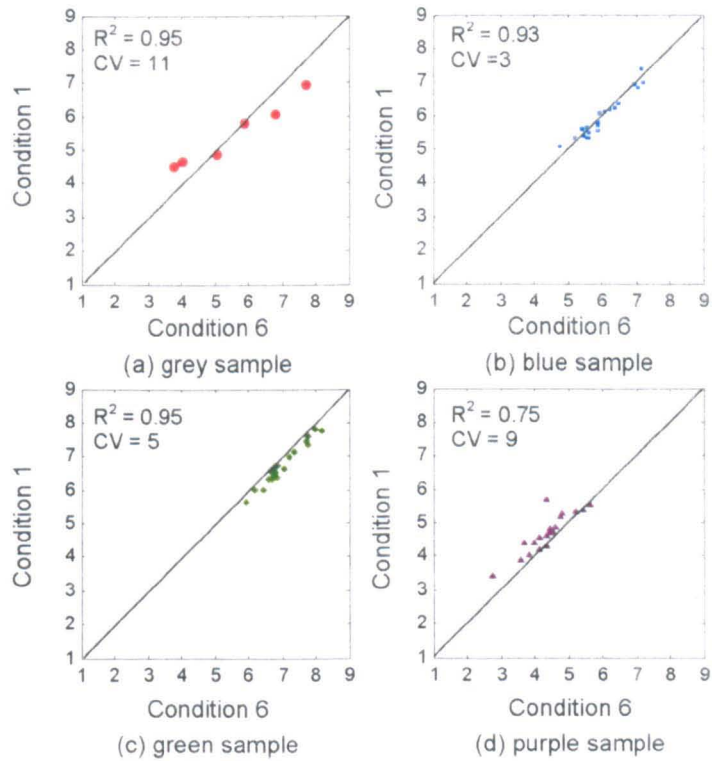


Figure 4-16: Comparisons of the scale values of the samples in each colour group between Condition 6 and Condition 1.

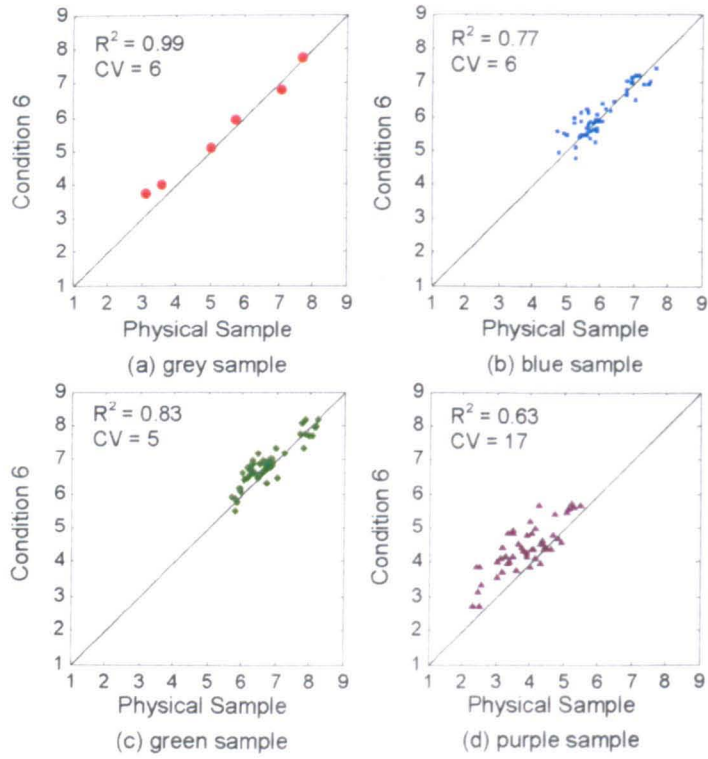


Figure 4-17: Comparisons of the scale values of the samples in each colour group between Condition 6 and the physical samples.

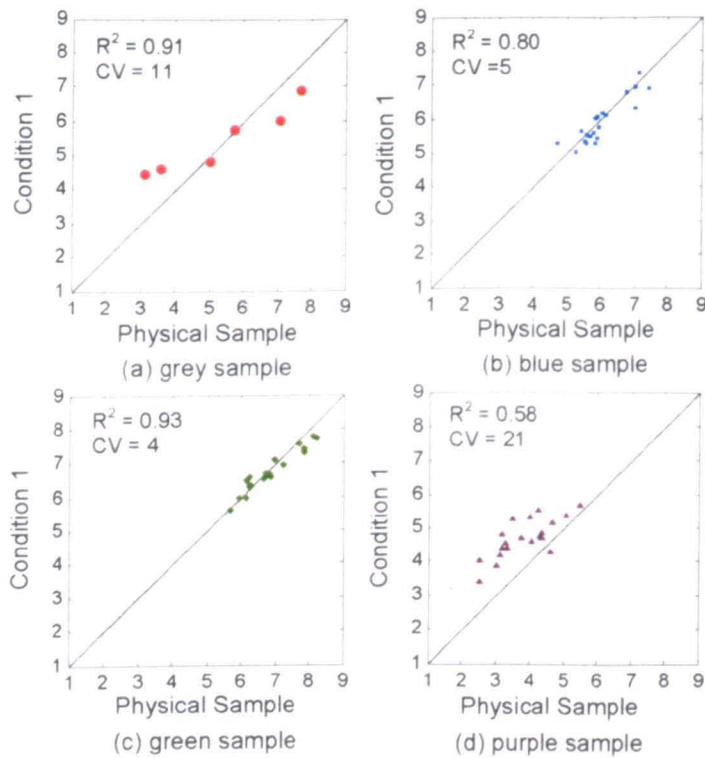


Figure 4-18: Comparisons of the scale values of the samples in each colour group between Condition 1 and the physical samples.

4.6.3 Comparisons of Perceptual Coarseness and Coarseness Model Predictions

In the previous Chapter 3, the coarseness model performance was verified by comparing the model predictions with the perceptual coarseness of the physical samples. The model predicted the coarseness from the images of the physical samples captured with high resolution and using the experimental viewing conditions. Hence, the model considers the high resolution image (832×832 pixel) as a physical sample size of 8×8 cm which could not be simulated on the display because of the limited display resolution. Therefore, here, the model performance was evaluated for the images and the conditions under which the observers really assessed the displayed image samples (Condition 6 and Condition 1). Because of errors that occurred during the image generation process, the reproduced image samples varied colorimetrically in a small way. However, the model should be able to predict the coarseness of the reproduced image samples under the conditions that those image samples were seen by the observers.

The scale values obtained in the previous sections were used to test the performance of the model for Condition 6 and Condition 1. The computed model predictions for Condition 6 and Condition 1 are all given in Appendix II. The model predictions were compared with the perceptual coarseness scale values for Condition 6, Condition 1 and also for the physical samples, together with indications of the associated correlation values in terms of R^2 and CV in Figure 4-19. This figure shows that the scale values agree well with the model predictions for all conditions with equivalent accuracy. The correlation between the model predictions and scale values had R^2 values of 0.91, 0.93 and 0.91 for Condition 6, Condition 1 and the physical samples respectively and CV values of 7, 10 and 7 for each three condition respectively. However, as well as the results of the physical samples, there were slight deviations from a linear relationship at low values of coarseness. In order to see these details, the model predictions and the scale values were plotted separately for each colour of the samples and each of the three conditions in Figure 4-20. Black dashed lines in the four graphs on the left column (in column (a)) indicate a linear regression line between the model predictions and the scale values of all samples obtained in Condition 6. Similarly, black dashed lines in the four graphs in each column (b) and (c) are linear regression lines for the data sets of all the samples in each of Condition 1 and the physical samples. Red lines in each graph are an indication of the linear regression line for each plotted

data set (not all the samples). The values of slope α and intercept β for all the regression lines, and the correlation measures, R^2 and CV, are given in Table 4-12. In this experiment, a unit of the perceptual coarseness (a 1-9 scale) and a unit of the model prediction are not the same. Therefore, the absolute value of the slopes is not important and the intercept does not necessary to be zero. The difference between those values indicates the tendency of the data sets.

A similar tendency was found for each colour group in the three experimental conditions except for the results for the grey colour samples in Condition 1 as shown in Figure 4-20. The model predicted the perceptual coarseness of the grey colour samples with the equivalent accuracy for all three conditions in terms of R^2 (0.98, 0.97 and 0.96 for Condition 6, Condition 1 and the physical samples respectively), while there was a dispersion from the majority of the samples (indicated with the black dashed line in Figure 4-20) at low coarseness levels in Condition 6 and the physical samples, however, not in Condition 1. This can also be seen from the differences in the slope and intercept between all the samples and the grey samples (Table 4-12). However, it is not possible to identify the reason, because the number of the grey colour samples was too small.

In all three conditions, the results for the purple samples also showed deviation from a linear regression line. A possible reason is for the deviation that the model underestimates the perceptual coarseness at the low coarseness levels. However, at this stage, this can not be proved because the observer variability was poor for the purple colour samples and therefore the errors might be caused by the observers' results not the underestimations by the model. Another possible reason is that the model is not capable of predicting the samples with particular colours such as purple, since it only considers the luminance information in the images. It is also difficult to investigate, since no other colour samples had the values of coarseness at such a low levels. Only limited sample colours with only limited coarseness distribution were available for this experiment. Therefore, the further study is suggested for the investigation of the influence of the chromatic channels for the model using samples with a wider range of colours and coarseness levels, although the use of only luminance channel to predict coarseness was thought as reasonable according to the correlation between the perceptual coarseness scaled using the image samples and the physical samples (see Section 4.6.2).

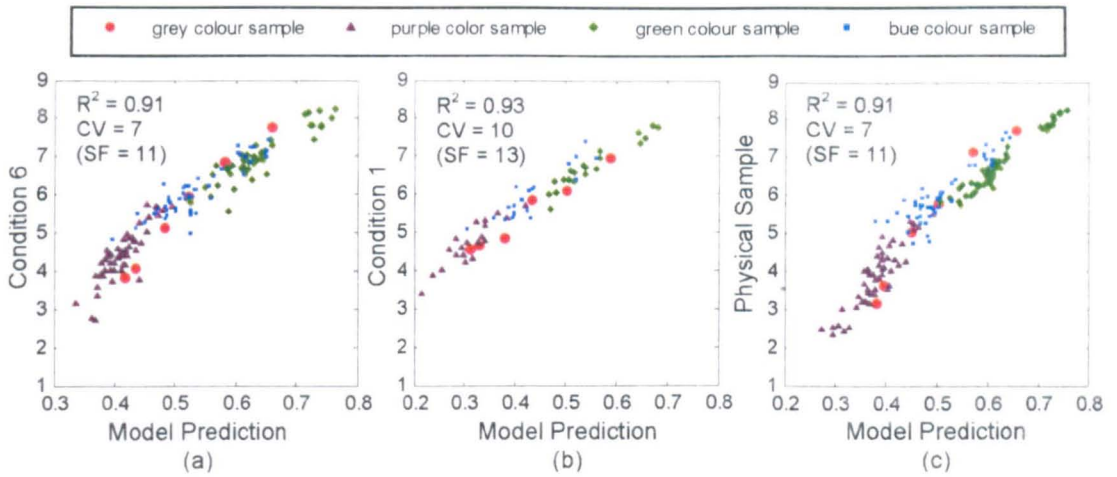


Figure 4-19: The scale values (a) in Condition 6 using the image samples, (b) in Condition 1 using the image samples (c) in the viewing cabinet using the physical samples, plotted as a function of the associated model predictions.

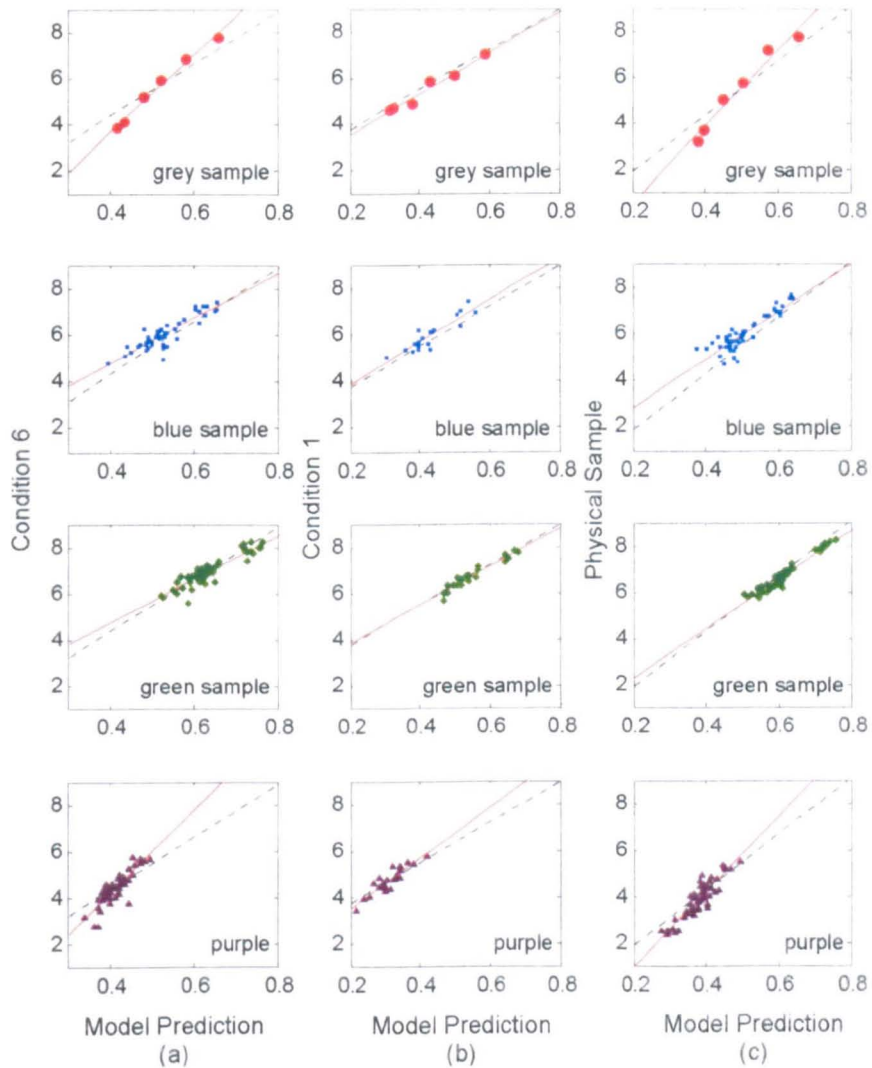


Figure 4-20: Scale values for samples in individual colour groups in (a) Condition 6, (b) Condition 1, and (c) the viewing cabinet using physical samples plotted against their respective model predictions. Black dashed lines are linear regression lines (the scale values against the model predictions) for all samples in each condition. Red lines indicate linear regression lines for the actual plotted data.

Table 4-12: Measures of model performance in terms of R^2 and CV with the scaling factor SF (see Section 2.10.3) and the values of the slope α and the intercept β for a linear regression line fitted to data (the scale values against the model predictions).

	All	Grey	Blue	Green	Purple
Condition 6					
α	11.25	16.89	9.54	9.40	17.95
β	-0.15	-3.17	0.97	0.95	-3.02
R^2	0.91	0.98	0.79	0.83	0.75
CV	7	10	5	4	10
SF	11	11	11	11	11
Condition 1					
α	8.69	8.96	9.04	8.37	10.84
β	1.99	1.65	2.02	2.09	1.29
R^2	0.93	0.97	0.82	0.92	0.81
CV	10	7	7	5	7
SF	13	13	14	12	15
Physical Sample					
α	12.12	16.70	10.31	10.70	16.27
β	-0.58	-2.87	0.74	0.10	-2.35
R^2	0.91	0.96	0.79	0.95	0.81
CV	7	13	6	3	12
SF	11	11	12	11	10

4.7 Summary

The investigations in this chapter were attempted to verify the images of the metallic-coating samples (physical samples) which were captured using a camera and were used as input for the coarseness model developed in Chapter 3, because the camera resolution was insufficient to capture individual aluminium flakes which were the major cause of the non-uniform appearance of the metallic coatings. In order to examine whether adequate information relevant to the perceptual coarseness was captured, the images were displayed on an LCD and their coarseness appearance was visually assessed. In addition to this verification, visual assessments using the images explored the possibility of using a digital image for appearance assessment instead of physical samples. This would provide an extremely convenient method of assessment for product/quality control, for the development of products and for communication.

To carry out the visual assessments, the images captured by the digital camera were converted to be suitable for displaying on the LCD. Although the accuracy of the camera and display characterisation processes was high (Section 3.4.2.1 and 4.3.3), the colour shifts were found between the physical samples and the displayed image samples.

The lightness of the images was relatively preserved, however the significant shifts were found in chroma and hue (Figure 4-11). These colour differences probably originated in the limitations of the LCD to display the fine details. Therefore, the absolute appearance of coarseness of the physical samples could not be reproduced using the displayed images. Another difficulty when using the images for visual assessments was that, although the resolution of the camera was not high enough to resolve individual aluminium flakes, the resolution of the display was even lower. Consequently, the actual physical size of the image sample had to be larger than that of the physical sample in order to display the whole area of the physical sample. Therefore, it was impossible to perform the visual assessment using the images under similar experimental conditions used to assess the physical samples in the viewing cabinet (see Section 3.3.2.1). Therefore, the perceptual coarseness of the image samples was scaled in two optimal conditions.

In one condition, the image of the whole area of the physical sample was displayed and the viewing geometry was reproduced in proportion to that used in the assessments for the physical samples. Thus, the viewing distance was greater than that used for the physical samples experiment, but the angular display size was kept at a similar value. The perceptual coarseness in this condition found to be smaller than that observed on the real samples. In the second condition, only a part of the image of the physical sample was displayed, but the angular display size was kept at a similar value, and the physical image size and the viewing distance were similar to the conditions used for scaling the physical samples. It was found that the perceptual coarseness in this condition appeared closer to that observed on the real samples.

However, the overall observer variability obtained in those two conditions for image samples and also in the viewing cabinet using the physical samples (see Section 3.3) was similar. This indicates that the observers assessed the coarseness with the equivalent accuracy under these three conditions, although the absolute appearances of the coarseness under these conditions were different. The overall agreement of the relative scaled perceptual coarseness between these three conditions was also encouraging. This indicates that the influences of chroma and hue to the perceptual coarseness were small. Also this supports the coarseness model to use only the luminance channel to predict perceptual coarseness. Moreover, these results indicates that the resolution of the images used in this experiment was high enough for the

coarseness appearance reproduction, even though each pixel size of the camera setting was larger than the size of the aluminium flakes.

In this chapter, the model was also implemented based on the image samples and the specification of the experimental conditions as input for each condition. The model predictions for each condition agreed well with the scale values. As well as the agreement of the scale values between three conditions, this also verified the quality of the reproduced images in terms of the perceptual coarseness. Moreover, it proved the performance of the coarseness model.

Although this study succeeded to visualise the relative coarseness appearance, but the further study is necessary to reproduce accurate colours and absolute coarseness for the use of images for total appearance assessment.

Chapter 5

Assessing and Modelling Glint using Physical Samples

5.1 Introduction

This chapter describes a study aimed to characterise perceptual glint of metallic-coating panels. As well as coarseness introduced in Chapter 3 and 4, “glint” has been identified as an important attribute of visual texture of metallic coatings by our collaborator, Akzo Nobel (2004-2006). Glint is as an attribute of visual texture and it is categorised as micro appearance, not as macro appearance such as gloss or specular reflection (see Section 2.5.1). Glint is originated mainly in characteristics of aluminium flakes contained in coatings. Obviously, the amount and orientation of flakes are assumed to be strongly related to perceptual glint. However, as well as the coarseness study presented in Chapter 3, this study intended to characterise perceptual glint focusing on what can be observed on the surface and is not concerned about physical components of metallic coatings.

As has been mentioned in Section 2.5.1.2, sparkle, brilliance and glitter are also used to describe similar or same appearance to glint and it remains unclear whether these terms refer exactly to the same phenomenon. Therefore, it is important to give a clear definition for the term “glint” as visual texture of metallic coatings together with a specification of observation conditions.

Several definitions were proposed by researchers in Akzo Nobel (2004) in order to describe glint of metallic coatings. Three of them are given:

- Points of reflected light of very high intensity that switch on and off while changing panel orientation.
- The impression that coatings show bright tiny lights under specific viewing angles only when irradiated by an intense directed light source.
- Tiny spot that is strikingly brighter than its surrounding. It is visible under directional illumination conditions only. The glint may be expected to switch on and off when the observation geometry is changed.

A final definition of glint was settled as given below by taking into account these proposed definitions.

“Tiny spot that is strikingly brighter than its surrounding, in other words, bright sparkle. It is visible under directional illumination conditions only. The glint may be expected to switch on and off when the illumination and observation geometry is changed.”

In respect to the definition given above, the characterisation of the perceptual glint of metallic-coating panels was carried out. The following sections start with an investigation of illumination and viewing geometry appropriate for observation of the glint of metallic coatings. Then, visual assessments to quantify perceptual glint using a series of metallic-coating panels are described. Finally, a model was developed to predict the perceptual glint of metallic coatings utilising a high dynamic-range (HDR) image of the metallic-coating panels (see Section 2.7). Hence, a procedure of the model development and its performance are then introduced.

5.2 Visual Assessments of Glint

The following sections introduce a method to characterise glint perceptually. Appropriate illumination and viewing geometry to observe glint and a method to quantify glint of metallic-coating panels are discussed as well as results thus obtained.

5.2.1 Samples

A set of 106 metallic-coating panels produced by Akzo Nobel were used as samples. These panels were made by mixing solid-colour pigments and aluminium flakes in different proportions. There were 6 grey colour panels, 20 blue colour panels, 20 brown colour panels, 20 green colour panels, 20 red colour panels and 20 yellow colour panels. The 6 grey colour panels had the same proportion of solid-colour pigments, but various amounts of metallic flakes, in order to produce a range of degrees of perceptual glint. As well as the grey colour panels, 20 panels of each colour group had the same proportion of solid-colour pigments with different amounts of aluminium flakes. In this experiment, a part of a panel (size of 6 × 6 cm) was presented to observers. Note that the 6 grey colour samples were the same as the grey colour samples used in the coarseness experiment in Chapter 3, but the 100 coloured samples

were different from those that were used in the coarseness experiment. Images of one sample from each colour group are given in Appendix III as examples. Note that the appearance of the images is not the same as that perceived for the physical samples because some of pixel values in those images were saturated.

5.2.2 Viewing Conditions

The glint, as it has been defined in the previous section, is visible under directional illumination. It becomes less visible when the illumination becomes less directional. It disappears under diffuse illumination conditions, but another attribute, coarseness, appears instead. It is known that the glint appears only under directional illumination, but the specification of illumination and viewing geometry which optimally enhances appearance of the glint is unknown. It can be considered that it is appropriate and easier for observers to characterise the perceptual glint under conditions for which the glint is most obvious. Therefore, illumination and viewing geometry conditions were first investigated. To do this, visual assessments were carried out using a tilting table so that the glint was observed at various illumination and viewing angles, and the angle where the glint pronounced most strongly, was determined by observers. The details of the experimental conditions and procedure are given in the following sections.

5.2.2.1 Experimental Settings

A schematic diagram of the experimental settings is given in Figure 5-1. Visual assessments were conducted in a darkened room. Each observer was seated at a desk, on which a chin-rest, a tilting table and a lamp were situated. Observers' viewing geometry was kept constant by adjusting the height of his/her eyes at 46 cm from the desk using the chin-rest. On the tilting table, two samples (metallic-coating panels), a reference and a test sample, were placed on the uniform mid-grey background. The tilting table allowed observers to see the glint effect of the samples at various illumination and viewing geometries by changing the angle of the tilting table. The angle of the tilting table was always measured with a digital protractor placed on the tilting table. A GretagMacbeth Sol-source lamp consisted of a tungsten halogen light source was arranged directly above the centre of the viewing field (Figure 5-1 (b)).

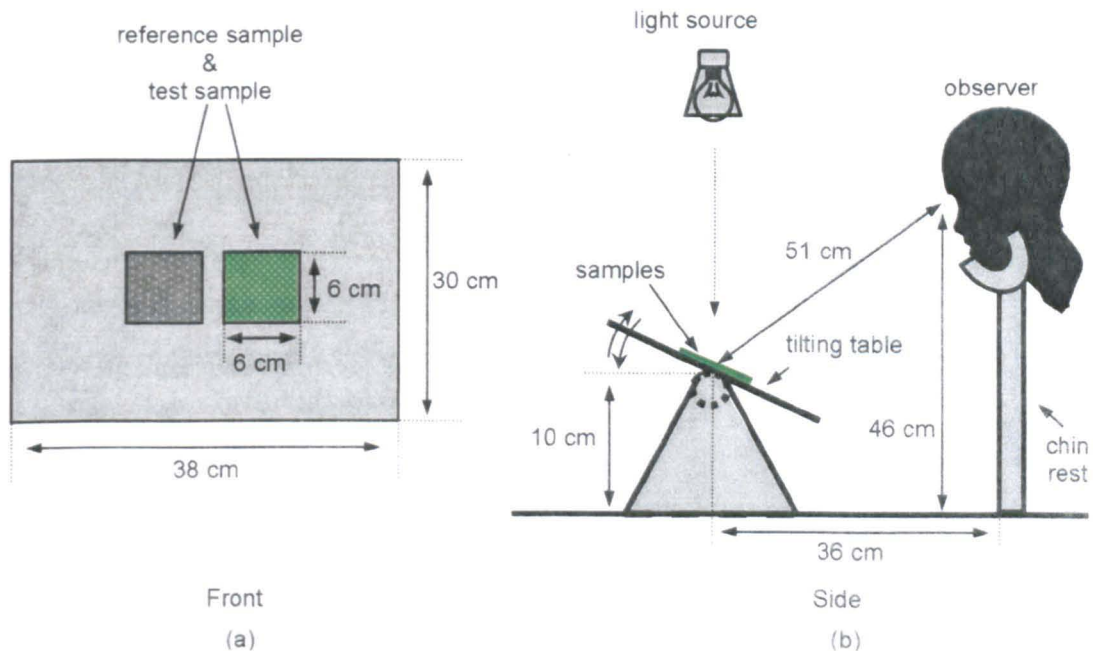


Figure 5-1(a) The samples' arrangements presented to observers. (b) Experimental settings.

The SPD of the light source was specified by measuring the SPD of the light reflected from a white ceramic standard tile using a Minolta CS1000 tele-spectroradiometer (TSR, see Section 3.2.1) in a 45/0 geometry. Figure 5-2 (a) shows the measurement condition and the measured SPD is given in Figure 5-2 (b). The CIE (x, y) chromaticity coordinates of the light source were equal to (0.4086, 0.4097) in terms of the CIE 1964 standard colorimetric observer. The uniform mid-grey background of the samples had CIELAB (L^* , a^* , b^*) values equal to (50.65, -2.018, -2.609) according to the measurement using 45/0 geometry at the centre of the viewing field. Spatial uniformity of the illumination was evaluated by measuring the viewing field at 15 various points using a GretagMacbeth LightSpex spectroradiometer on the flat table as shown in Figure 5-3; (a) for the measurement condition; (b) for the locations where the measurements were made; and (c) for the illuminance in lux at the locations corresponding to Figure 5-3 (b). It is ideal to use a light source having directional illumination which is strong enough to bring out glint appearance and is able to illuminate a viewing field uniformly. A large high-power light would be ideal, as would sunlight, however, because of the limited availability of light sources for this experiment, the GretagMacbeth Sol-source lamp was employed, although it was not possible to illuminate the viewing field uniformly. However, the illuminance at the centre of both samples (a reference and a test sample) was adjusted to be as close as possible.

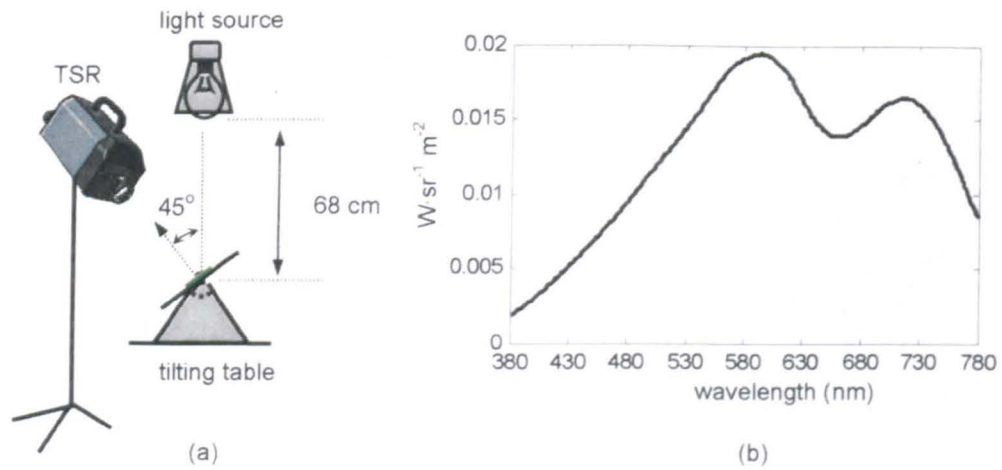


Figure 5-2: (a) Illumination and viewing geometry for the measurement. (b) SPD of the light source.

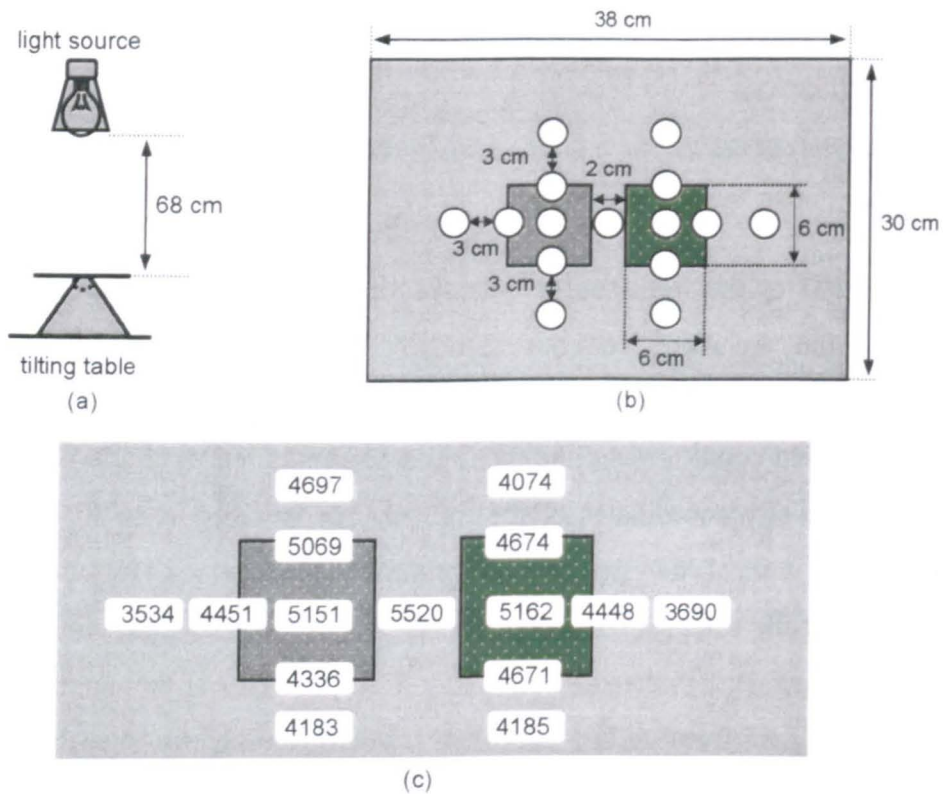


Figure 5-3: Spatial uniformity evaluation of the illumination. (a) Geometry of measurements in uniformity evaluation. (b) Locations of measurements in uniformity evaluation. (c) Illuminance variation in lux at the 15 points.

5.2.3 Scaling the Perceptual Glint

Ten observers, including five females and five males aged between 25 and 38, participated in the visual assessment for scaling the perceptual glint of the metallic-coating samples. All observers were either students or members of staff at the University of Leeds. They all passed the Ishihara visual test as observers with normal colour vision.

Magnitude estimation method (see Section 2.9.2) was applied as a scaling method. Each observer commenced an observation session by adapting to the viewing field which was the mid-grey background illuminated by the lamp in a darkened room. Before the assessment starts, the following instruction was given to each observer:

Your task is to assess “glint” of pairs of metallic-coating panels. The ‘glint’ has the following definition:

“Tiny spot that is strikingly brighter than its surrounding, in other words, bright sparkle. It is visible under directional illumination conditions only. The glint may be expected to switch on and off when the observation geometry is changed. Glint value can be defined by local contrasts between “bright sparkle” and its surround, and the amount of the sparkle.”

In each assessment, two metallic-coating panels will be provided, a test sample and a reference sample. Both are placed on a tilting table whose angle can be adjusted. Please find an angle that provides you with the maximum perception of glint based on the reference sample and then assign a number that best describes the perception of glint of the test sample in comparison with the reference sample having a glint value of 50. A value of 1 represents no glint at all. This is an open-ended scale, i.e., there is no upper limit to this scale.

One of the six grey samples, having a middle level of glint, was chosen as a reference sample. The perceptual glint of the remaining 105 samples was visually scaled by comparison with the reference sample. All samples were presented in a random order and the position of the reference sample was also randomly selected on either the right- or left-hand side of the test sample during the assessment so as to prevent any systematic and experimenter bias (Dean & Voss, 1999). In each assessment (*i.e.*, every time when a new test was presented to the observer), the angle of the tilting table was changed by an experimenter in order to let the observer re-adjust the angle best to see the glint. This was to examine the consistency of angle selections within observers and between observers. Moreover, this was to compensate for any slight variations in the observer’s position which were not preventable during a long session (each session lasted about one hour).

At each assessment, a scaled glint value and the selected angle of the tilting table were recorded. A training session was given before the main experimental session to allow the observers to practice and understand the glint observed on the metallic-coating samples. In the training session, each observer was asked to find an angle that provides the maximum perceptual glint of each of the six grey colour samples. After the observer got used to participating in the assessment, the main session involving the assessments of all the 105 assessments was carried out. Each observer conducted the main session twice, in order to test repeatability. A total of 20 main sessions corresponding to a total of 2100 magnitude estimations were made by 10 observers (10 observers \times 2 sessions \times 105 samples).

5.2.4 Data Analysis and Results

5.2.4.1 Scale Value

For the perceptual glint scaling, the magnitude estimation method with the reference sample was applied. A standard way to summarise results is to take a geometric mean over the results from all observers for each sample to obtain a mean scale value (see Section 2.9.2), since each observer may use different scales. To see the influence of scale difference, arithmetic and geometric means of all raw observer data were calculated for each sample and Figure 5-4 investigates their correlation by plotting the arithmetic means against the geometric means with a line which indicates perfect agreement between them. Their correlation of an R^2 value of 0.998 indicates very good agreement. However, it can be clearly seen in Figure 5-4 that the correlation gradually declines toward the largest scaled values of perceptual glint. This might be caused by positively skewed distributions of the observer data or rather by a few high outlying values. A range of the numbers used in all the sessions to scale the glint was from 1 to 400, while the maximum numbers used at each session were various. The maximum values were in between 80 to 200 in 12 sessions and between 201 to 300 in 7 sessions out of 20 sessions. This is not surprising, since the glint was scaled with an open-ended method. However, the correlation between the arithmetic and geometric means was good; namely there was not much scale variation in this experiment. This is presumably because the reference sample was used to compare with the test samples and therefore, the observers used a relatively similar range of the numbers. However, some small influence of the scale differences can still be observed. Consequently, the geometric

means were found to be appropriate to determine the mean observer scale values of the perceptual glint for the samples in this study. The derived geometric means for all the samples used in this study were given in Appendix IV.

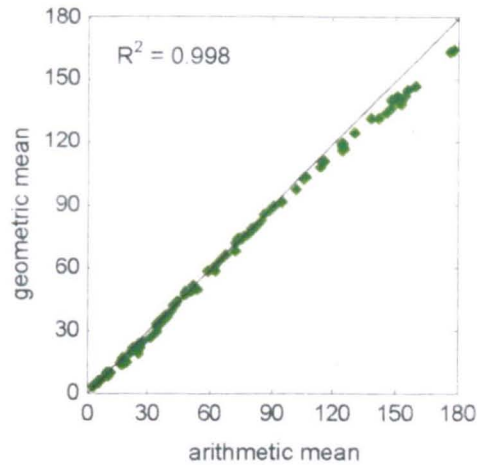


Figure 5-4: Comparison of the scale values derived from geometric mean and arithmetic mean of raw observer data for all samples.

5.2.4.2 Observer Variability

In this section, reliability of the visual assessments was investigated by analysing observer variability.

Individual raw observer data are often examined in terms of a relation to geometric means of all observer data in a logarithmic form as shown in Equation 5-1 (Bartleson, 1979; Luo and Clarke *et al.*, 1991; Pointer, 1980).

$$\log \bar{O}_i = \alpha_j \log O_{ij} + \beta_j \quad \text{Equation 5-1}$$

where O_{ij} is the raw data of the observer j for the test sample i , \bar{O}_i is the geometric mean of the test sample i , and the α_j and β_j are a slope and an intercept of a regression line determined using least-squares fitting for the observer j . This function (Equation 5-1) is not always appropriate to describe a relation between an individual observer data and a geometric mean of all observer data. This function is appropriate, however, if individual observer data are either logarithmic or power functions of stimuli (Bartleson, 1979); these are the functions which are the most likely when human sensation is scaled using magnitude estimation (Luce & Krumhansl, 1988). Unfortunately, there is no evidence in the relation between the observer data and the stimuli in this experiment, since there is no certain physical measure for the glint; namely stimuli. However, this function was found to be suitable to describe the relationship from the good correlations

between the logarithmic observer data and geometric means. Table 5-1 shows that the correlation for all the samples is 0.92 on average with a range from 0.86 to 0.96 in terms of R^2 . These high correlations indicate that a log-log scale is appropriate to analyse observer variability; by comparison, the correlation of the raw observer data and the geometric mean is 0.89 on average with a range from 0.75 to 0.95 (see Table 5-1). Therefore, logarithms of the data were utilised to investigate observer accuracy and repeatability (see Section 3.3.4.2).

Observer accuracy was investigated in terms of R^2 and CV based on a logarithm scale, and also R^2 values in a raw scale in order to compare with that in logarithm scale. The slope α_j and intercept β_j in Equation 5-1 was also obtained to see the differences of scales used in each session from the mean observer data. The slope and intercept in Table 5-1 show absolute differences of the slope and the intercept from the mean observer data. It means that the differences from the value of one for the slope ($| \alpha_j - 1 |$) and from the values of zero for intercept ($| \beta_j - 0 |$) were calculated, because the slope of one and the intercept of zero indicate the mean observer data. Table 5-1 summarises these observer accuracy measures in terms of their mean, median, maximum and minimum of all the sessions from all the samples and from individual grey, blue, brown, green, red and yellow samples. Table 5-3 to Table 5-9 give the accuracy measures of R^2 and CV for individual session from all the samples and individual coloured sample groups together with the raw slope value α_j and the raw intercept value β_j .

Similarly, Table 5-2 summarises mean, median, maximum and minimum observer repeatability measures of all the observers from all the samples and individual coloured sample groups. It shows R^2 and CV values between the observer's first and second session in the logarithm scale and R^2 values between them in the raw scale. In addition, the details of repeatability are given in Table 5-10 to Table 5-16.

Table 5-1: A summary of observer accuracy measures from all the samples and the samples in each grey, blue, brown, green, red and yellow sample group.

Sample	All Samples	Grey Samples	Blue Samples	Brown Samples	Green Samples	Red Samples	Yellow Samples
Logarithmic Scale							
Slope ($\alpha - 1$)							
Mean	0.20	0.24	0.22	0.24	0.20	0.19	0.19
Median	0.18	0.19	0.18	0.24	0.20	0.20	0.17
Max	0.38	0.98	0.49	0.63	0.49	0.37	0.61
Min	0.00	0.00	0.03	0.00	0.00	0.02	0.01
Intercept ($\beta - 0$)							
Mean	0.35	0.33	0.37	0.41	0.35	0.35	0.33
Median	0.33	0.30	0.35	0.38	0.36	0.31	0.25
Max	0.68	0.61	0.78	1.12	0.69	0.82	0.96
Min	0.02	0.04	0.02	0.03	0.03	0.02	0.02
R²							
Mean	0.92	0.92	0.94	0.93	0.93	0.93	0.93
Median	0.93	0.93	0.95	0.94	0.95	0.94	0.94
Max	0.96	0.96	0.98	0.98	0.98	0.98	0.97
Min	0.86	0.83	0.88	0.87	0.78	0.78	0.84
CV							
Mean	14	9	12	14	13	14	16
Median	13	9	11	13	13	14	14
Max	23	23	22	27	22	25	31
Min	7	4	6	6	6	6	8
SF (for CV)							
Mean	0.98	0.99	0.99	0.98	0.99	0.98	0.98
Median	0.98	0.98	0.98	0.98	0.99	0.97	0.97
Max	1.12	1.05	1.12	1.11	1.06	1.17	1.17
Min	0.91	0.93	0.91	0.89	0.91	0.9	0.88
Raw Scale							
R²							
Mean	0.89	0.93	0.91	0.92	0.92	0.91	0.92
Median	0.91	0.94	0.92	0.92	0.94	0.91	0.92
Max	0.95	1.00	0.98	0.96	0.97	0.97	0.99
Min	0.75	0.83	0.78	0.87	0.78	0.81	0.82

Table 5-2: A summary of observer repeatability measures from all the samples and the samples in each grey, blue, brown, green, red and yellow sample group.

Sample	All Samples	Grey Samples	Blue Samples	Brown Samples	Green Samples	Red Samples	Yellow Samples
Logarithmic Scale							
R²							
Mean	0.93	0.91	0.88	0.87	0.90	0.89	0.93
Median	0.97	0.92	0.87	0.88	0.91	0.90	0.97
Max	0.99	0.95	0.95	0.97	0.93	0.97	0.99
Min	0.73	0.83	0.79	0.71	0.83	0.78	0.73
CV							
Mean	7	13	17	15	16	17	7
Median	6	13	16	15	18	16	6
Max	14	17	25	22	21	30	14
Min	1	8	10	10	11	11	1
Raw Scale							
R²							
Mean	0.95	0.87	0.85	0.89	0.89	0.88	0.95
Median	0.96	0.88	0.85	0.91	0.92	0.90	0.96
Max	0.99	0.96	0.94	0.95	0.96	0.97	0.99
Min	0.84	0.71	0.73	0.76	0.77	0.78	0.84

Table 5-3: Observer accuracy measures for each session from all the samples.

Observer	Session	Logarithmic Scale					Raw Scale
		Slope (α)	Intercept (β)	R ²	CV	SF (for CV)	R ²
1	1	0.90	0.11	0.96	8	0.97	0.94
	2	1.07	-0.22	0.95	9	0.94	0.92
2	1	0.86	0.28	0.91	13	1.02	0.88
	2	1.00	0.04	0.96	7	1.02	0.92
3	1	1.24	-0.44	0.86	16	0.98	0.86
	2	1.38	-0.68	0.91	15	0.98	0.84
4	1	1.17	-0.45	0.88	15	0.92	0.92
	2	0.82	0.26	0.92	12	0.97	0.93
5	1	0.67	0.58	0.94	21	0.99	0.75
	2	0.67	0.59	0.93	22	1.01	0.86
6	1	0.88	0.14	0.93	10	0.95	0.88
	2	0.85	0.11	0.95	9	0.91	0.89
7	1	1.17	-0.33	0.94	11	0.98	0.92
	2	0.82	0.32	0.94	12	1.01	0.95
8	1	0.68	0.58	0.96	21	1.00	0.91
	2	0.69	0.61	0.92	23	1.04	0.83
9	1	0.94	-0.02	0.94	9	0.93	0.95
	2	1.19	-0.51	0.93	13	0.91	0.93
10	1	1.38	-0.55	0.86	16	1.03	0.90
	2	1.20	-0.12	0.88	13	1.12	0.89

Table 5-4: Observer accuracy measures for each session from the grey colour samples.

Observer	Session	Logarithmic Scale					Raw Scale
		Slope (α)	Intercept (β)	R ²	CV	SF (for CV)	R ²
1	1	1.30	0.08	0.96	6	0.97	0.99
	2	1.06	-0.21	0.96	6	0.97	1.00
2	1	0.87	0.24	0.90	6	1.05	0.88
	2	1.00	0.04	0.96	5	1.04	0.96
3	1	1.98	-0.29	0.84	10	1.01	0.96
	2	1.34	-0.61	0.92	10	1.04	0.88
4	1	1.10	-0.31	0.87	10	0.97	0.94
	2	0.80	0.27	0.92	10	1.05	0.94
5	1	0.66	0.60	0.92	12	0.93	0.83
	2	0.68	0.61	0.93	10	0.96	0.85
6	1	0.87	0.13	0.94	4	0.98	1.00
	2	0.85	0.10	0.95	4	0.95	0.99
7	1	1.19	-0.37	0.93	7	0.99	0.88
	2	0.82	0.33	0.94	5	0.97	0.98
8	1	0.69	0.57	0.96	14	0.97	0.99
	2	0.71	0.58	0.94	23	0.96	0.90
9	1	0.95	-0.05	0.94	12	1.00	0.90
	2	1.19	-0.51	0.93	8	0.98	0.98
10	1	1.40	-0.58	0.83	8	1.03	0.91
	2	1.19	-0.09	0.87	10	1.05	0.92

Table 5-5: Observer accuracy measures for each session from the blue colour samples.

Observer	Session	Logarithmic Scale					Raw Scale
		Slope (α)	Intercept (β)	R ²	CV	SF (for CV)	R ²
1	1	0.89	0.14	0.97	7	0.96	0.98
	2	1.19	-0.39	0.96	9	0.96	0.95
2	1	0.97	0.08	0.94	8	1.02	0.89
	2	1.03	-0.02	0.97	6	1.02	0.94
3	1	1.49	-0.78	0.92	15	1.01	0.87
	2	1.35	-0.56	0.92	13	1.01	0.89
4	1	1.34	-0.70	0.93	13	0.94	0.92
	2	0.88	0.22	0.95	10	1.00	0.96
5	1	0.68	0.54	0.97	17	0.96	0.78
	2	0.63	0.65	0.95	22	0.97	0.86
6	1	0.87	0.14	0.88	12	0.94	0.92
	2	0.86	0.10	0.96	7	0.91	0.86
7	1	1.14	-0.27	0.98	7	0.98	0.93
	2	0.84	0.30	0.95	11	1.02	0.95
8	1	0.66	0.61	0.96	19	0.98	0.95
	2	0.68	0.62	0.93	20	1.03	0.81
9	1	0.92	0.06	0.96	7	0.95	0.96
	2	1.16	-0.46	0.95	11	0.91	0.93
10	1	1.40	-0.58	0.88	14	1.04	0.90
	2	1.19	-0.12	0.89	11	1.12	0.87

Table 5-6: Observer accuracy measures for each session from the brown colour samples.

Observer	Session	Logarithmic Scale					Raw Scale
		Slope (α)	Intercept (β)	R ²	CV	SF (for CV)	R ²
1	1	0.86	0.20	0.96	9	0.98	0.96
	2	0.98	-0.10	0.92	11	0.93	0.91
2	1	0.76	0.47	0.94	18	1.04	0.91
	2	1.00	0.04	0.97	6	1.03	0.89
3	1	1.44	-0.77	0.92	16	0.97	0.87
	2	1.63	-1.12	0.94	18	0.95	0.87
4	1	1.35	-0.80	0.91	16	0.90	0.96
	2	0.80	0.27	0.92	13	0.95	0.95
5	1	0.72	0.52	0.98	19	1.02	0.88
	2	0.73	0.49	0.95	18	1.02	0.87
6	1	0.87	0.16	0.95	9	0.96	0.91
	2	0.82	0.18	0.93	11	0.92	0.92
7	1	1.17	-0.29	0.93	11	0.99	0.95
	2	0.83	0.32	0.95	12	1.02	0.93
8	1	0.68	0.56	0.97	20	0.99	0.93
	2	0.66	0.66	0.87	27	1.06	0.88
9	1	0.93	-0.03	0.93	10	0.92	0.96
	2	1.24	-0.64	0.91	15	0.89	0.89
10	1	1.32	-0.44	0.94	12	1.04	0.95
	2	1.24	-0.18	0.90	12	1.11	0.95

Table 5-7: Observer accuracy measures for each session from the green colour samples.

Observer	Session	Logarithmic Scale					Raw Scale
		Slope (α)	Intercept (β)	R ²	CV	SF (for CV)	R ²
1	1	0.93	0.08	0.97	6	0.98	0.96
	2	1.08	-0.23	0.96	9	0.95	0.93
2	1	0.95	0.11	0.88	13	1.02	0.83
	2	1.00	0.07	0.97	7	1.04	0.91
3	1	1.14	-0.24	0.78	18	1.00	0.86
	2	1.30	-0.51	0.94	12	0.99	0.87
4	1	1.07	-0.21	0.85	15	0.95	0.96
	2	0.79	0.36	0.95	13	0.99	0.97
5	1	0.71	0.45	0.91	18	0.95	0.78
	2	0.72	0.52	0.97	18	1.01	0.92
6	1	0.87	0.16	0.96	8	0.96	0.93
	2	0.81	0.22	0.96	9	0.93	0.94
7	1	1.23	-0.46	0.97	10	0.96	0.97
	2	0.81	0.36	0.94	14	1.02	0.97
8	1	0.68	0.57	0.98	20	0.98	0.94
	2	0.70	0.60	0.95	22	1.04	0.81
9	1	0.92	0.03	0.91	11	0.93	0.96
	2	1.29	-0.69	0.94	14	0.91	0.96
10	1	1.49	-0.67	0.93	14	1.06	0.96
	2	1.31	-0.40	0.96	12	1.06	0.93

Table 5-8: Observer accuracy measures for each session from the red colour samples.

Observer	Session	Logarithmic Scale					Raw Scale
		Slope (α)	Intercept (β)	R ²	CV	SF (for CV)	R ²
1	1	0.88	0.13	0.95	9	0.95	0.91
	2	1.05	-0.16	0.98	7	0.95	0.92
2	1	0.77	0.44	0.93	17	1.04	0.87
	2	1.03	-0.04	0.96	8	1.01	0.91
3	1	1.17	-0.34	0.86	16	0.96	0.91
	2	1.28	-0.51	0.91	15	0.97	0.81
4	1	1.37	-0.82	0.92	17	0.90	0.94
	2	0.85	0.17	0.92	12	0.95	0.97
5	1	0.68	0.62	0.97	25	1.04	0.94
	2	0.68	0.59	0.92	24	1.02	0.85
6	1	0.86	0.21	0.94	11	0.98	0.91
	2	0.96	-0.08	0.96	7	0.91	0.94
7	1	1.28	-0.58	0.93	15	0.94	0.91
	2	0.88	0.18	0.95	10	0.98	0.96
8	1	0.69	0.60	0.96	24	1.03	0.87
	2	0.71	0.55	0.95	21	1.03	0.84
9	1	0.98	-0.12	0.98	6	0.91	0.97
	2	1.21	-0.54	0.94	13	0.90	0.97
10	1	1.22	-0.27	0.78	18	1.04	0.88
	2	1.18	-0.02	0.91	12	1.17	0.94

Table 5-9: Observer accuracy measures for each session from the yellow colour samples.

Observer	Session	Logarithmic Scale					Raw Scale
		Slope (α)	Intercept (β)	R ²	CV	SF (for CV)	R ²
1	1	0.94	0.03	0.96	8	0.96	0.93
	2	1.06	-0.24	0.95	11	0.92	0.91
2	1	0.97	0.04	0.95	10	0.99	0.94
	2	0.97	0.08	0.96	9	1.02	0.93
3	1	1.12	-0.29	0.87	16	0.94	0.87
	2	1.44	-0.82	0.92	19	0.94	0.82
4	1	0.99	-0.15	0.90	14	0.90	0.91
	2	0.79	0.25	0.93	14	0.92	0.95
5	1	0.65	0.65	0.91	30	1.02	0.92
	2	0.66	0.67	0.91	31	1.05	0.92
6	1	0.91	0.02	0.95	10	0.92	0.89
	2	0.82	0.12	0.96	10	0.88	0.94
7	1	1.16	-0.24	0.95	12	1.01	0.96
	2	0.80	0.40	0.97	16	1.04	0.99
8	1	0.72	0.55	0.94	24	1.04	0.88
	2	0.73	0.57	0.93	25	1.08	0.91
9	1	0.96	-0.09	0.94	10	0.91	0.96
	2	1.11	-0.36	0.92	14	0.90	0.95
10	1	1.61	-0.96	0.85	22	0.99	0.90
	2	1.27	-0.14	0.84	17	1.17	0.89

Table 5-10: Observer repeatability measures of each observer from all the samples.

Observer	Logarithmic Scale		Raw Scale
	R ²	CV	R ²
1	0.92	13	0.92
2	0.90	14	0.89
3	0.81	11	0.84
4	0.88	19	0.91
5	0.90	18	0.82
6	0.90	14	0.83
7	0.91	18	0.92
8	0.92	18	0.80
9	0.89	14	0.88
10	0.80	16	0.87

Table 5-11: Observer repeatability measures of each observer from the grey colour samples.

Observer	Logarithmic Scale		Raw Scale
	R ²	CV	R ²
1	0.99	1	0.99
2	0.94	6	0.95
3	0.98	5	0.98
4	0.73	14	0.84
5	0.85	11	0.99
6	0.99	5	0.99
7	0.94	5	0.94
8	0.98	9	0.95
9	0.99	9	0.96
10	0.97	3	0.96

Table 5-12: Observer repeatability measures of each observer from the blue colour samples.

Observer	Logarithmic Scale		Raw Scale
	R ²	CV	R ²
1	0.93	13	0.96
2	0.95	8	0.88
3	0.90	8	0.87
4	0.93	17	0.90
5	0.91	16	0.90
6	0.89	13	0.71
7	0.91	15	0.95
8	0.95	13	0.80
9	0.92	13	0.85
10	0.83	14	0.85

Table 5-13: Observer repeatability measures of each observer from the brown samples.

Observer	Logarithmic Scale		Raw Scale
	R ²	CV	R ²
1	0.86	18	0.87
2	0.93	20	0.87
3	0.81	10	0.83
4	0.95	20	0.94
5	0.95	12	0.73
6	0.89	14	0.81
7	0.86	20	0.89
8	0.85	25	0.82
9	0.89	15	0.83
10	0.79	15	0.90

Table 5-14: Observer repeatability measures of each observer from the green colour samples.

Observer	Logarithmic Scale		Raw Scale
	R ²	CV	R ²
1	0.94	11	0.92
2	0.87	14	0.91
3	0.71	16	0.81
4	0.83	22	0.95
5	0.92	17	0.76
6	0.90	14	0.92
7	0.88	21	0.95
8	0.97	12	0.87
9	0.82	18	0.89
10	0.88	10	0.94

Table 5-15: Observer repeatability measures of each observer from the red colour samples.

Observer	Logarithmic Scale		Raw Scale
	R ²	CV	R ²
1	0.93	12	0.94
2	0.92	21	0.91
3	0.85	11	0.81
4	0.90	19	0.93
5	0.93	18	0.82
6	0.93	16	0.94
7	0.91	18	0.92
8	0.91	19	0.77
9	0.91	12	0.96
10	0.83	18	0.92

Table 5-16: Observer repeatability measures of each observer from the yellow colour samples.

Observer	Logarithmic Scale		Raw Scale
	R ²	CV	R ²
1	0.94	13	0.90
2	0.94	11	0.91
3	0.86	13	0.91
4	0.89	18	0.89
5	0.84	30	0.92
6	0.90	15	0.82
7	0.97	21	0.93
8	0.91	20	0.81
9	0.91	12	0.91
10	0.78	22	0.81

5.2.4.3 Glint Angle

The glint can be seen on a metallic coating at various geometries as long as it is viewed under directional illumination. However, intensity of the glint changes with viewing geometries. Therefore, illumination and viewing geometry which provides the maximum perceptual glint, was investigated. The observers adjusted the tilting table to the angle which provides the maximum perceptual glint, at each assessment.

The results of the selected angles are given in Figure 5-5. The angle θ used in this study is the angle with respect to the tilting table which is illustrated in Figure 5-5 (a). Figure 5-5 (b) shows the maximum, minimum and mean angles selected at each session. The horizontal axis indicates the observer and his/her session (*e.g.*, the result of Observer 1's first session is labelled as Ob1-1st) and the vertical axis indicates the angle. 95 % confidence intervals of each session are plotted with their mean in Figure 5-5 (c).

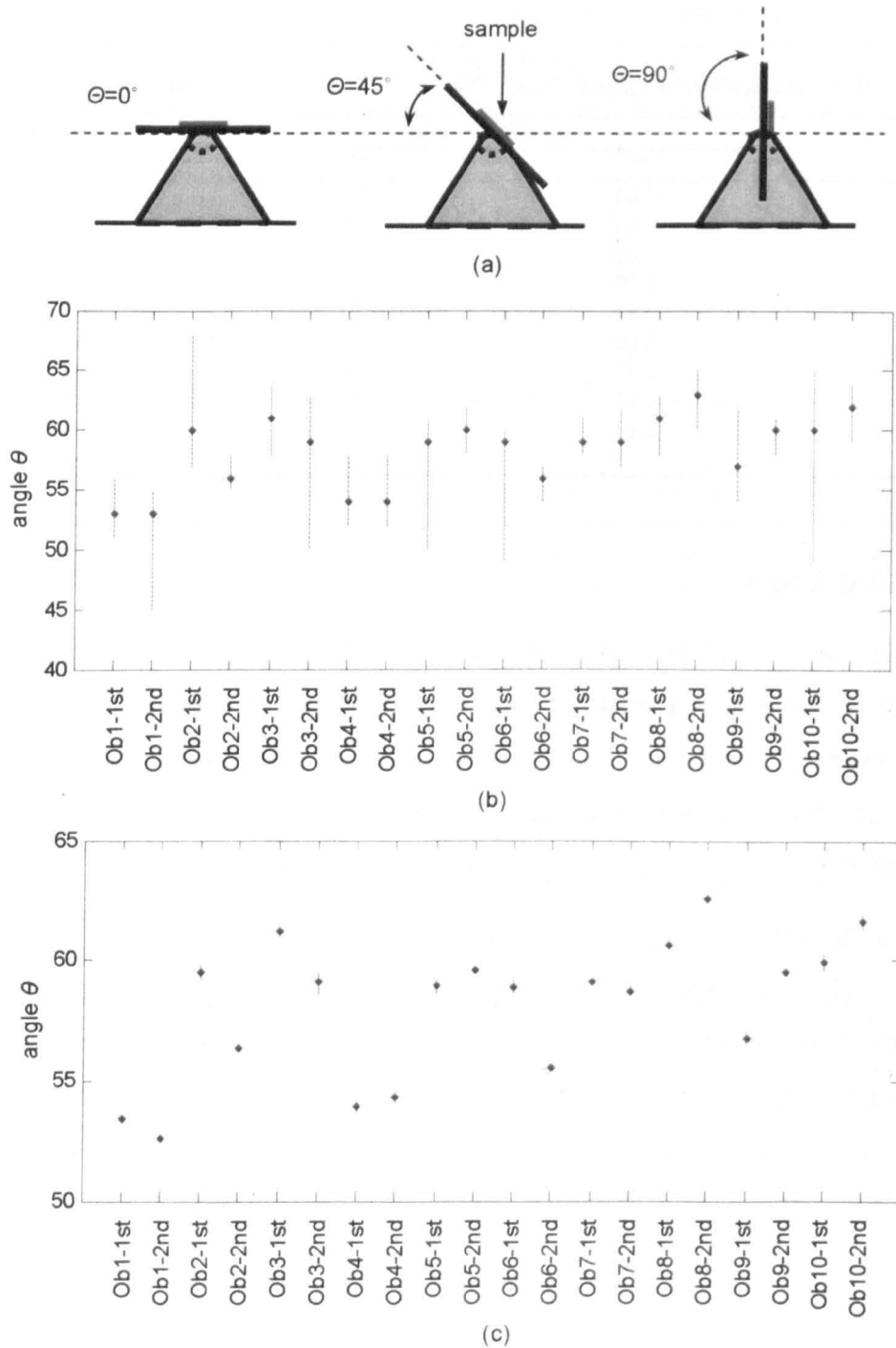


Figure 5-5: The results of the selected angles. (a) The indication of the angle measurement. (b) The maximum, minimum and mean angles selected at the each session (e.g., the result of Observer 1's first session is labelled as Ob1-1st). (c) 95 % confidence intervals of each session were plotted with their mean.

5.2.5 Discussion

Reliability of the experimental results was examined in terms of the observer accuracy and repeatability. The results of the accuracy can be found in Table 5-1 and Table 5-3 to Table 5-9. The results of the repeatability are in Table 5-2 and Table 5-10 to Table 5-16.

The mean accuracy was 0.92 from all the samples with a range from 0.86 to 0.96 in terms of R^2 . The mean R^2 values of 0.92, 0.94, 0.93, 0.93, 0.93 and 0.93 were obtained for the grey, blue, brown, green, red and yellow samples respectively. This indicates that individual observer data are linearly correlated with the mean observer data. It is also found that there is no significant difference in observer accuracy depending on the colour of the samples. The mean CV values of 14, 9, 12, 14, 13, 14 and 16 were obtained from all the samples, and the samples in each colour group respectively. The grey samples had a better accuracy in comparison with the other colours. It was thought that there were only five grey samples and they had relatively distinguishable perceptual glint differences, therefore the observers could judge consistently, while there were 20 for each in other colour groups. The CV values also proved that there were no differences in observer accuracy between the samples having different colours except the grey samples.

Since the magnitude estimation method was applied, it was not surprising to find variations in the scale used by observers; namely, the range of the numbers used for scaling varied between observers. However, it can be seen from the results of the slope in Table 5-1 and Table 5-3 to Table 5-9 that there were relatively small scale differences. The mean of the absolute differences in the slope and the intercept between the sessions were only 0.20 and 0.35 respectively from all the samples. This indicates that the majority of the observers used quite similar scales. However, it is also evident that a minority of the observers used quite different ranges of numbers from the others. A large slope difference obtained from a session for all the samples was 1.38. Although this indicates the large scale difference from the one which the majority used, the accuracy of this session was the R^2 value of 0.91 which was as accurate as the other sessions. Thus it should be noted that the slope is only the indication of variations of the scale used by the observers and it does not need to be same for all sessions. In contrast to slope, the intercepts should be a value of close to zero. This is because the bottom end value of one was given as an indication of no glint in the visual assessment,

while there was no indication for top end in a glint scale, since the open-ended system was used. Even a largest intercept difference from the mean observer data was 0.68 from all the samples. This is considerably small comparing with the range of scale values used by the observers (see Section 5.2.4.1).

The results of the observer repeatability can be found in Table 5-2 and the details in Table 5-10 to Table 5-16. As well as the accuracy, the repeatability for the grey samples was slightly superior to that for the other coloured samples and it is because of the number of the samples and their distinguishable glint levels. However the performance of the repeatability also does not seem to depend on the colours of the samples.

Another study was also carried out, in parallel with the present study, to investigate glint appearance of coating materials by Akzo Nobel (Kirchner and Kieboom *et al.*, 2007). The definition of glint in their study (see Section 2.5.1.2) made much the same point as the definition given in this study. Seven observers assessed a total of 216 samples. These samples were not only metallic-coating panels but also pearlescent-coating panels and some panels coated with a mixture of metallic and pearlescent. Similar to the coarseness experiment by Akzo Nobel (see Section 3.3.5), a 0-9 categorical judgement scaling method was applied with eight references which corresponded to categories one to eight and they allowed the observers to assign the values up to a half scale precision. As in this study, directional illumination was utilised, but the glint was scaled at three fixed illumination and viewing angles with a fixed angle for the reference samples. The angles were 35°, 45° and 60° (note that the angle was indicated in the same matter to the present study see Figure 5-5 (a)). The viewing conditions used by Akzo Nobel are given in Figure 5-6.

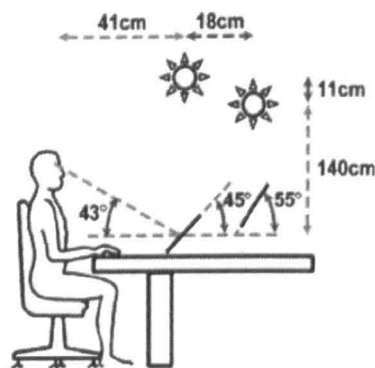


Figure 5-6: Sketch of the experimental set-up used at Akzo Nobel (Kirchner and Kieboom *et al.*, 2007). An example of the angle of 45°.

They obtained the average observer accuracy of 0.94, 0.90 and 0.94 at the 35°, 45° and 60° conditions respectively in term of R^2 . It found to be very similar to the result of the present study which was 0.92. Their results of accuracy and repeatability indicate that the glint was able to be assessed equivalently at the three different geometries. However, the maximum glint values obtained in three conditions were varied. It was smaller at the 60° condition (the glint value rounded off to 8) than the other two conditions. They explained that the perceptual glint decreased when the observation angle was further away from the specular angle. In fact, the 35° condition had largest maximum values, of which was the condition closest to a specular angle among three conditions followed by the 45° condition. However, in contrast, this study reports a mean angle of 58° (see Figure 5-5), *i.e.*, the aspecular angle was 71°. The mean angle range was from 53° to 63° and the maximum and minimum of the all the observations' angles were 45° and 68° respectively (Figure 5-5) from the results of the present study. It tells that although the range of the angles that the observers selected, was not so small, no observer selected less than 45°. This disagreement can be considered that the observers scaled the glint using slightly different criteria. The definitions of the glint in both experiments were similar. However, in the present study, the observers were guided that:

“Glint value can be defined by local contrasts between “bright sparkle” and its surround, and the amount of the sparkle.”

As a result of this instruction, it can be assumed that observers in this study scaled the glint in terms of the difference between the bright sparkle and its surround (local contrast), while the Akzo Nobel observers, the glint was evaluated in simply in terms of the brightness of the sparkle (absolute intensity). When the samples are observed close to the specular angle, not only the sparkles but also the whole area of the samples look brighter compared with when the samples are viewed away from the specular angle, but the contrast between bright sparkles and their surround dramatically decreases and even some of the tiny spots are vanished. It is suggested that this is the reason why in this study the optimal angle for glints was somewhat different to that found in the Akzo Nobel experiment.

However, the results obtained by Akzo Nobel indicate that the glint was assessed with equivalent accuracy at any of the three angles, although there were differences in the scaled glint value between the observation angles. This suggests, in any event, that in this study the variations found in observation angles between the observers were not

critical (Figure 5-5). The lack of importance of the viewing angle in this study may result because the observers always assessed the test samples relative to the reference sample (and both samples were always at the same angle). In the Akzo Nobel experiment, the angle for the reference samples was fixed and the test samples were assessed at all three angles which resulted in variations in the scaled glint values found between angles.

5.3 Computational Model for Perceptual Glint Prediction

The following sections describe a model developed for predicting the perceptual glint using a digital colour image of metallic coatings. Unlike coarseness, the presence of glint means that it is impossible to capture the full dynamic range of such a scene in a single image due to the limitations in most image-capture devices. Hence, an HDR image system was employed (see Section 2.7) and a glint model was developed based on the HDR image as input. A framework of the model is given in Figure 5-7. The following sections describe each stage of these processes. They start with a procedure for creating the HDR image followed by various steps to extract a feature correlated with perceptual glint. The performance of this model is then evaluated.

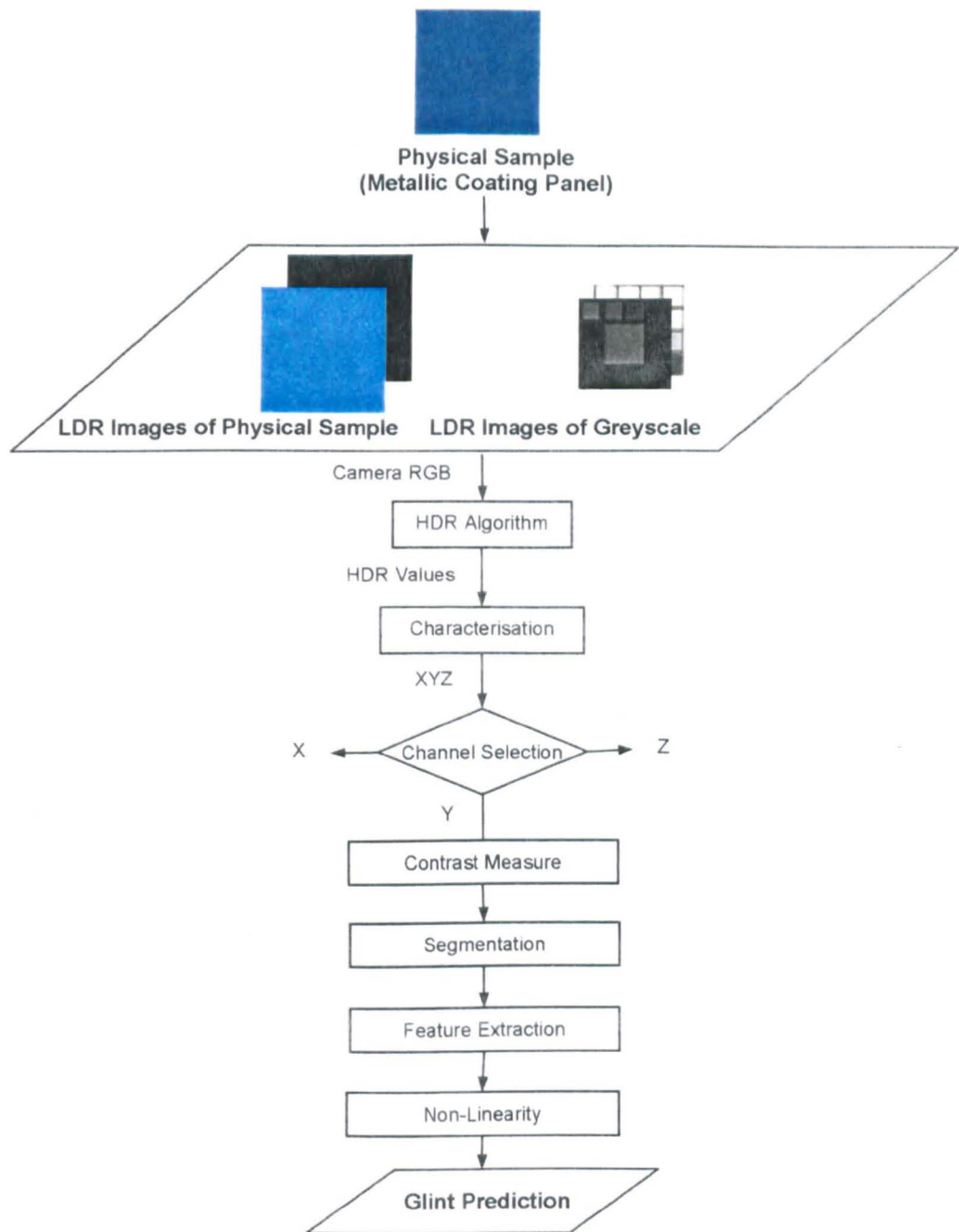


Figure 5-7: A flowchart of the main framework for the glint model

5.3.1 Measurement and Capture of Targets

In order to develop and evaluate a model, digital images and spectral properties, such as the SPD and the CIE XYZ tristimulus values of the 106 metallic-coating samples (same samples as used in the visual assessment in Section 5.2) and a GretagMacbeth ColorChecker DC (ColorChecker DC) were required. The measurement of the spectral properties and the image capturing were carried out in consistent conditions with the visual assessments (see Figure 5-1 (b)). The TSR or a digital camera was placed where the observer's eyes would be as shown in Figure 5-8.

The angle of the tilting table was adjusted at $\theta = 58^\circ$ (aspecular angle of 71°) which was the average angle of all the visual assessments (see Section 5.2.4.3). Each target was placed at the centre of the tilting table.

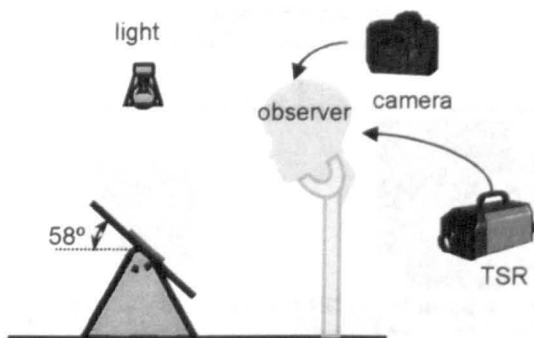


Figure 5-8: An illustration of the experimental condition for the spectral properties measurements using a TSR and the image capturing using a digital camera.

The SPD of the metallic-coating samples and 12 patches in a greyscale on the centre region of the ColorChecker DC were measured using the TSR. From the obtained SPD, CIE XYZ tristimulus values were computed using the CIE 1964 standard colorimetric observer. However, it was difficult to carry out measurements for all patches in the ColorChecker DC on the tilting table, because of the limitation of the space on the tilting table and the structure of the ColorChecker DC (the patches stuck on a board). Hence, spectral reflectance of all patches in the ColorChecker DC was measured using the CE7000A with specular component excluded (see Section 3.2.2). CIE XYZ tristimulus values were then computed using the SPD of the illumination obtained by measuring the white patch in the greyscale and the CIE 1964 standard colorimetric observer.

Digital images of the samples, the 12 greyscale patches and the ColorChecker DC were captured using a Nikon D1X digital camera (see Section 3.4.1) with two different exposure settings.

5.3.2 HDR Image Capture

If a dynamic range of a scene is greater than that of an image-capture device then a problem occurs that is illustrated in Figure 5-9 (a) and (b). Although the camera gain can be altered, at a camera setting, the camera response saturates well before the maximum scene intensity is reached (Figure 5-9 (a)); at a different setting, the converse problem occurs where the camera captures the high-intensity end of the scene, but

provides little useful information at the bottom end (Figure 5-9 (b)). This situation occurs when an image of scene is captured containing metallic-coating panels with glint because bright spots, which are mainly caused by the reflection from aluminium flakes in coatings, are extremely brighter than their surround. Conventional cameras are not able to capture both those bright and dark part of the scene intensity at once because of limitations in technology. Therefore, in this study, an HDR image was created in order to capture all the necessary information in the scene.

Each sample was captured using two different exposures and these two LDR (low dynamic-range) images were then combined to create an image with a higher dynamic-range than either of the original images. For example, a LDR image captured with an exposure setting (Exposure 1) is resulted in Figure 5-9 (a); another LDR image captured with a different exposure setting (Exposure 2) is resulted in Figure 5-9 (b); and these two LDR can be combined as illustrated in Figure 5-9 (c): the region A where the response of Exposure 1 can be used to recover scene properties; the region C where the response of Exposure 2 can be used; and the region B where the responses of both exposures can be used. A result of combining the data from Exposures 1 and 2 is seen in Figure 5-9 (d) where a virtual camera response is possible that exceeds the dynamic range of either of the individual exposures. This process generates an HDR image. Examples of images corresponding to Figure 5-9 (a) and (b) are given in Figure 5-10. Images in Figure 5-10 (a) capture a dark part of the scene and images in Figure 5-10 (b) have information of a bright part of the scene.

The problem of capturing HDR images from multiple exposures has been studied and a number of solutions have been proposed (Debevec & Malik, 1997; Krawczyk and Goesele *et al.*, 2005; Mitsunaga & Nayar, 1999; Nayar & Mitsunaga, 2000; Robertson and Borman *et al.*, 2003; Xiao and DiCarlo *et al.*, 2002). The published solutions treat each channel separately and attempt to derive smooth camera response functions. However, the problem that they are trying to address is generally more complex than the problem in this study. Typically, multiple-exposure images are taken in, for example, outdoor scene including bright part and shadow, or indoor scene with a window in daytime such as in a church, where there are no surfaces of known spectral properties, and illumination is also unknown and spatially variable. Since this study is aiming to capture images under the controlled condition of a laboratory, the camera response function was derived from the relationship between the camera responses and the

spectral properties of the greyscale. The details of the HDR capturing procedure are described in the following sections.

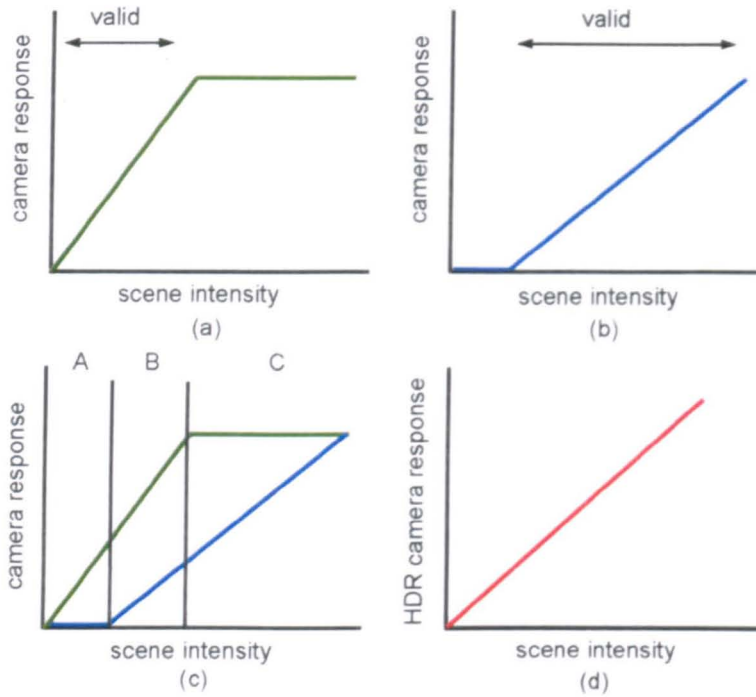


Figure 5-9: (a) Response of camera at Exposure 1. (b) Response of camera at Exposure 2. (c) Illustration of regions A (Exposure 1 valid only), B (both exposures valid) and C (Exposure 2 valid only). (d) A combined HDR camera response.

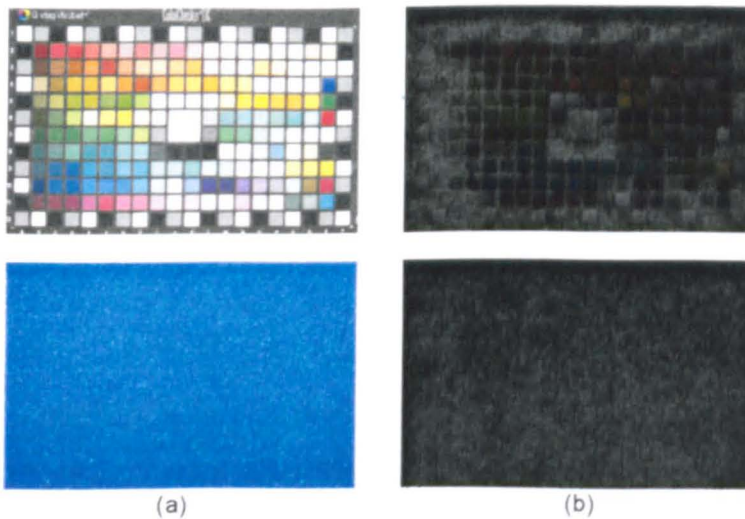


Figure 5-10: Examples of LDR images captured with two exposure setting. (a) Images capture information of a dark part of the scene (the right hand side image). (b) Images capture a bright part of the scene (the left hand side images).

5.3.2.1 LCR Image Capture

HDR images were created based on the images captured at two exposures and the spectral properties of the 12 greyscale patches including very dark (black) and bright (white) colour patches. Two exposure settings, 1/10 seconds (Exposure 1) and 1/400 seconds (Exposure 2), were selected resulting from a comparison of the camera responses at the various exposure settings of 1/2, 1/3, 1/4, 1/5, 1/6, 1/8, 1/10, 1/13, 1/200, 1/250, 1/320, 1/400, 1/400 and 1/500. In order to use information of the greyscale, camera responses of all greyscale patches should be within a valid range of a camera response at either one of the two exposure settings. As illustrated in Figure 5-11 (Exposure Setting Selection 1), if there are any patches not covered with any of exposure settings, it results in missing parts of scene information in an HDR image; information of middle levels of greyscale patches is missing. On the other hand, as is illustrated in Figure 5-11 (Exposure Setting Selection 2), if many of patches lay on ranges of both exposure settings, there is no missing scene information, but only a limited range of information is included. Consequently, the exposure settings of 1/10 and 1/400 were applied, for which there were no invalid areas and no widely overlapped area so as to use the information from the greyscale effectively.

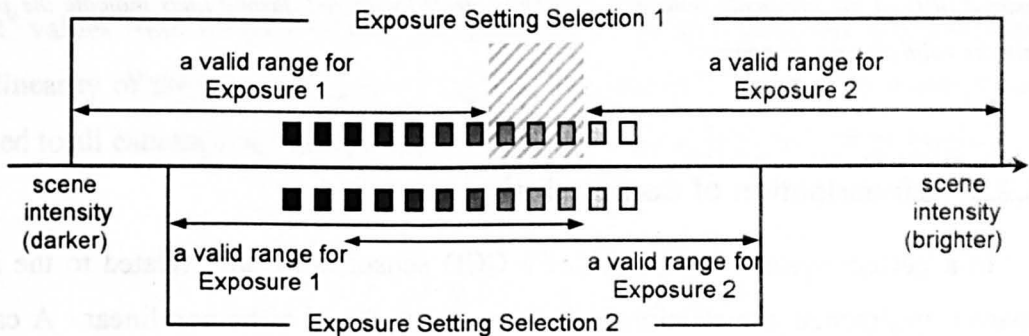


Figure 5-11: An example of selections of camera settings. If there are any patches are not covered with any of exposure settings, it results in missing parts of scene information (Exposure Setting Selection 1). On the other hand, there is no missing scene information, but only a limited range of information is included (Exposure Setting Selection 2).

Figure 5-12 shows the camera responses of the greyscale patches of the green channel from two of the selected exposure settings (normalised in the range 0-1) against the sum of the measured SPD of the corresponding greyscale patches. The plots circled with dashed lines indicate the patches within the valid camera responses. Although some of the un-circled patches were not under- or over-exposed (which means camera responses do not reach 0 or 1), these patches were excluded from consideration because

these camera responses at the close to under- and upper-limit are less reliable (Reinhard and Ward *et al.*, 2006). Note that the camera responses have not been linearised at this stage so that the plots in Figure 5-12 do not relate to exactly the same condition as Figure 5-9 (c) which assumed a perfect linear camera system. It would be possible to create an HDR image from more than two exposure settings, however it was assumed that two was sufficient to capture all the scene information needed in this study and it is verified later.

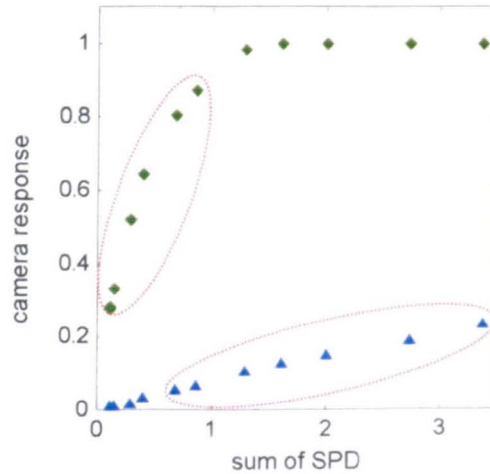


Figure 5-12: The camera responses (normalised in the range 0-1) of the green channel from two of the selected exposure settings (1/10 is shown in green and 1/400 is shown in blue) against the sum of the measured SPD of the greyscale patches. The plots circled with red dashed lines indicate the patches within the valid camera responses.

5.3.2.2 Linearisation of Camera Response

In a perfect system, the output of a CCD sensor is linearly related to the input. However, in practice, the relationship is sometimes found to be non-linear. A camera response is often approximately related to the output of a CCD with a power function as described by Equation 5-2.

$$dRGB = rRGB^{1/\gamma} \quad \text{Equation 5-2}$$

A camera response $dRGB$ is a raw response, $rRGB$, raised by an exponent $1/\gamma$. One reason for this non-linear response of many cameras is a manufacturer-induced correction for the non-linear response of typical display systems. The luminance L of typical display devices can be modelled as a function of input voltage V and an exponent γ which is usually referred to as the gamma of a display device as is given in Equation 5-3.

$$L = V^\gamma \quad \text{Equation 5-3}$$

The knowledge of the camera response function was required to combine images from two exposures into an HDR image. Therefore, the output of the camera was linearised with values that were linearly related with the camera input. For each red, green and blue channel, a power function was applied to fit the relationship between the valid camera responses and the sum of the SPD for the greyscale patches. Three graphs on the top of Figure 5-13 show the valid camera responses¹ of the greyscale patches plotted against the respective sum of the SPD with the fitted power function for the camera setting at Exposure 1. Three graphs on the bottom of Figure 5-13 show the linearised camera responses for the greyscale patches plotted against the sum of the SPD with the fitted line for each channel for the camera setting at Exposure 1. Figure 5-14 is similarly for the camera setting at Exposure 2. This transformation can be written with Equation 5-4.

$$IRGB = \alpha \times dRGB^\beta \quad \text{Equation 5-4}$$

where *IRGB* is each linearised red, green and blue camera response, *dRGB* is the camera responses and α and β are the scaling factor and the exponent respectively. The errors of the fitted lines were measured in terms of R^2 . The exponent, the scaling factor and the R^2 for each camera setting and each channel are given in Table 5-17. The range of the R^2 values from 0.986 to 0.998 indicates that the power functions could model the non-linearity of the camera responses well. These power functions were subsequently applied to all camera responses in order to linearise them before further processing.

¹ Only the valid camera responses were used for the linearisation. For Exposure 1 setting, camera responses equal to 255 were not included, and for Exposure 2 setting, camera responses equal to 0 were not included.

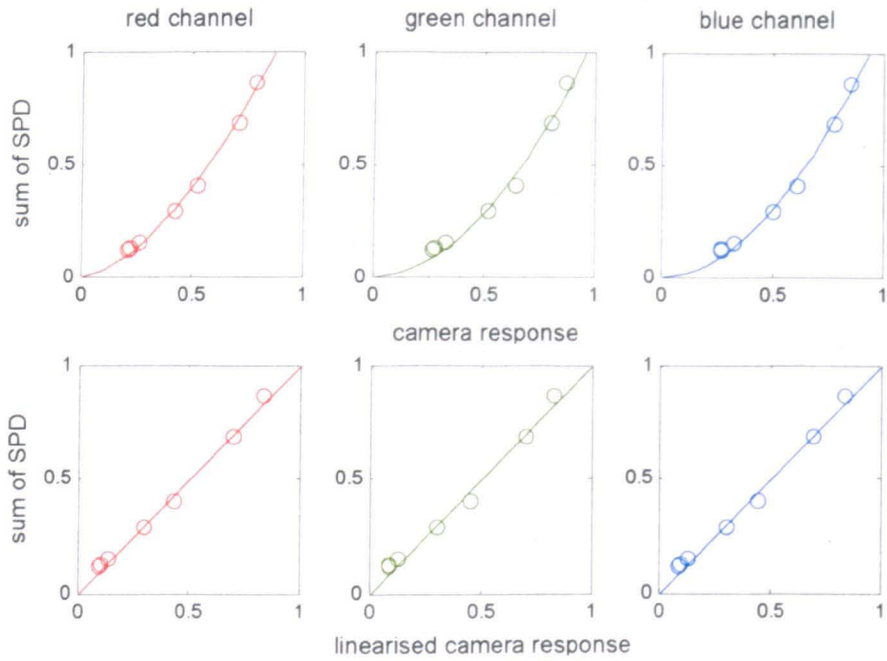


Figure 5-13: Camera responses of the valid patches of the greyscale for each red, green and blue channel plotted against the sum of the SPD of the corresponding patches (circles) with the fitted lines (solid lines) (graphs on the top row), and linearised camera responses converted using the power function (circles) with fitted lines (solid lines) (graphs on the bottom row) for the camera setting at the exposure time of $1/10$ (Exposure 1).

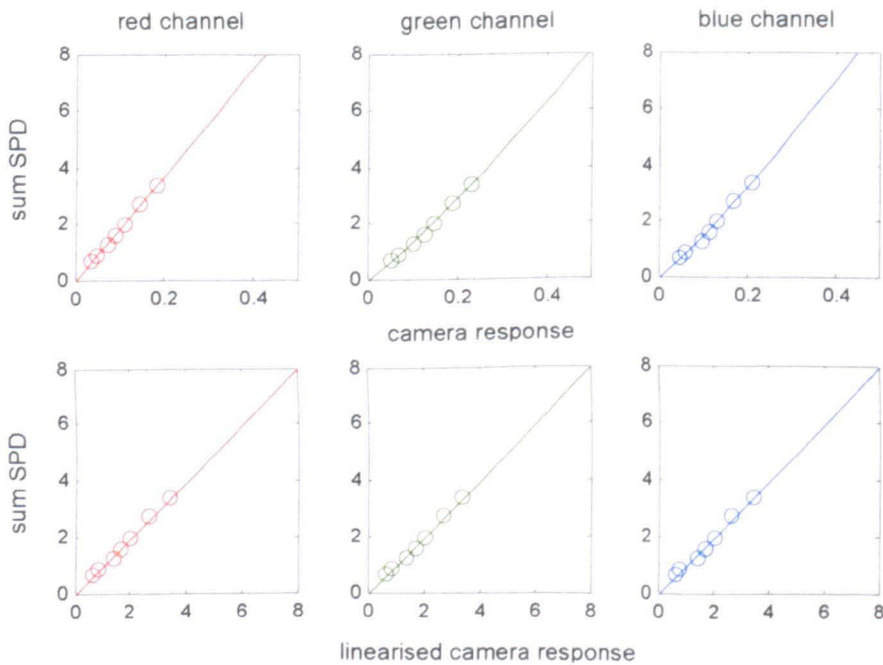


Figure 5-14: Camera responses of the valid patches of the greyscale for each red, green and blue channel plotted against the sum of the SPD of the corresponding patches (circles) with the fitted lines (solid lines) (graphs on the top row), and linearised camera responses converted using the power function (circles) with fitted lines (solid lines) (graphs on the bottom row) for the camera setting at the exposure time of $1/400$ (Exposure 2).

Table 5-17: The scaling factor α and the exponents β for the linearisation of the camera responses, and R^2 values as indications of how well the power functions fit to the greyscale patches.

Exposure Time	1/10			1/400		
	α	β	R^2	α	β	R^2
Red Channel	1.25	1.66	0.994	19.01	1.02	0.998
Green Channel	1.07	1.95	0.986	17.83	1.14	0.998
Blue Channel	1.14	1.94	0.990	19.97	1.14	0.993

5.3.2.3 Spatial Uniformity Correction

Spatial uniformity correction was performed to minimise the effect of non-uniformity of the intensity of the illumination, since the intensity of the illumination was not uniform over the capturing field due to the spot light used. Also, in the context of cameras, spatial uniformity correction compensates for any non-uniformity in sensitivity of individual elements of a CCD array and imperfections in an optical system; optical properties of most camera lenses allow more light to transmit at the centre area of lens than the peripheral area (Hong and Luo *et al.*, 2001) and chromatic aberration because of inevitable consequences of the laws of refraction at spherical surfaces. Hence, spatial uniformity correction was performed to the red, green and blue channels individually after the linearisation according to a previously established method (Hardeberg, 1999) as shown in Equation 5-5.

$$Q(i, j) = \frac{(IRGBw - IRGBd) \times (IRGB(i, j) - IRGBd(i, j))}{(IRGBw(i, j) - IRGBd(i, j))} \quad \text{Equation 5-5}$$

Equation 5-5 describes the relationship of a linearised camera response, $IRGB(i, j)$, of either red, green and blue channel at each pixel position (i, j) to its spatially-corrected value $Q(i, j)$. $IRGBw$ and $IRGBd$ are the mean linearised camera responses for central areas of the uniform white (bright) and black (dark) patches which fill the camera viewing field. $IRGBw(i, j)$ and $IRGBd(i, j)$ are the linearised camera responses for the white and black patches at each pixel position (i, j) respectively.

5.3.2.4 Deriving Camera Response Function

An HDR camera response function was derived by combining the two LDR camera responses (which were previously linearised and spatially corrected) so that the HRD response was correlated with the sum of the SPD of the greyscale patches. Figure 5-15 (a) and (b) show the two linearised LDR responses of the greyscale patches for the

green channel. Valid LDR camera responses from Exposure 1 were up to the upper end, $G_{1(u)}$, as indicated with A in Figure 5-15 (a) and the camera responses of Exposure 2 were valid above the lower end, $G_{2(l)}$, where the area indicated as C in Figure 5-15 (b). At the area B in Figure 5-15 (c), the average of the camera responses from the two exposures were used. The HDR camera responses of the greyscale patches are given in Figure 5-15 (d) for each red, green and blue channel.

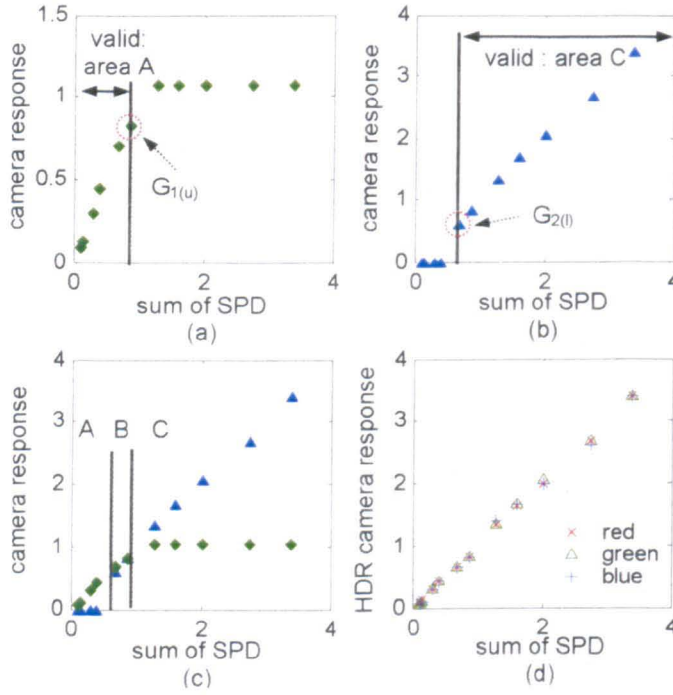


Figure 5-15: (a, b) Linearised camera responses against the sum of the SPD of the greyscale patches (green channel) for two exposures with the indications of the upper and lower end of the valid values. (c) The indications of the regions: A (Exposure 1 (exposure 1/10) valid only); B (both exposures valid); and C (Exposure 1 (exposure 1/400) valid only.) (corresponding to Figure 5-9(c)). (d) Normalised sum of the SPD against the HDR values combined from the two camera response of the greyscale patches (corresponding to Figure 5-9 (d)).

This procedure of combining the two LDR images to create an HDR image is summarised in Equation 5-6.

$$HDR(i, j) = \begin{cases} Q_1(i, j) & \text{if } Q_1(i, j) \leq G_{2(l)} \\ Q_2(i, j) & \text{if } Q_2(i, j) \geq G_{1(u)} \\ \frac{1}{2}(Q_1(i, j) + Q_2(i, j)) & \text{if } G_{2(l)} < Q_1(i, j) \text{ and } Q_2(i, j) < G_{1(u)} \end{cases} \quad \text{Equation 5-6}$$

where $HDR(i, j)$ indicates the HDR value of an image at a pixel position (i, j) , $Q_e(i, j)$ represents the LDR camera response of the exposure, e ($e = 1$ or 2 , since only two exposure settings were used in this study), after the linearisation and the spatial

correction were applied, and $G_{1(u)}$ and $G_{2(l)}$ are the responses of the greyscale at the upper end of Exposure 1 and the lower end of Exposure 2 respectively (see Figure 5-15 (a) and (b)). Figure 5-15 (d) allows a comparison between the actual sum of the SPD and the generated HDR values of the greyscale patches. For a perfect HDR response, the data in Figure 5-15 (d) would be expected to form a straight line at 45°. The errors between the two were CV of 1.29 for each of the red, green and blue channels and the root mean squared error (RMSE) values calculated after normalising the data in the range 0-1, were 0.010, 0.012 and 0.017 for the red, green and blue channels respectively. These results prove the performance of the HDR algorithm used in this study. The HDR images of all the samples (metallic-coating samples) were obtained from a set of LDR images using this algorithm. Using this imaging setup and HDR algorithm, it was possible to capture reflected light 5.6, 5.3 and 5.9 times greater than that from the white patch in the greyscale for the red, green and blue channels respectively. The generated HDR system was sufficient to cover the full range of the scene intensity of the samples; when all of the samples were considered, the maximum pixel value was about 69 %, 63 % and 43 % of the individual red, green and blue full HDR range available. This reveals that some of the pixels in the HDR images of the samples had the values several times greater than the similarly captured white patch and therefore justifies the use of the HDR approach in this study. The HDR algorithm used was relatively simple and based on the SPD of a greyscale. Only two exposure levels were used, but it would be possible to extend the method to combine three or more exposures using the linear method which was applied.

5.3.3 Colour Space Transformation

For the work in this chapter, the colorimetric response at each pixel in the image is required. The approach taken, is to capture an HDR image and then to use a characterisation method in order to convert the HDR values to device-independent values such as CIE XYZ tristimulus values. Similar to the work that has already been described (see Section 3.4.2.1), linear and polynomial regression models with least-squares fitting (see Section 2.8.3.1) were used for characterisation. Various training data sets and various terms of polynomial regression models were implemented and an appropriate model was chosen based on performances evaluated with a test data set consisting of mean HDR values for each of the HDR images of the 106 metallic-coating samples and the XYZ values of the corresponding samples from the measurements. The

augmented matrices used are given in Table 3-9 in Section 3.4.2.1. Three training data sets, which were combinations of the 106 metallic-coating samples and the 166 patches from in the ColorChecker DC (no glossy surface patches included), were compared. As has been mentioned, the XYZ values of the 106 metallic-coating samples were derived from the measurements of the spectral properties using the TSR at the condition illustrated in Figure 5-8 (see Section 5.3.1), while the XYZ values of the 166 patches in the ColorChecker DC were based on the measurement using the CE7000A (see Section 5.3.1). The spectral properties of surfaces often change with the viewing geometry. This change is large for gonioapparent materials, *e.g.*, metallic coatings, but smaller for matt surface materials, *e.g.*, a ColorChecker DC. The colour differences of the 12 greyscale patches in the ColorChecker DC between the measurements using the TSR and the measurements using the CE7000A were a mean ΔE^*_{ab} of 0.65 with a range from 0 to 1.22. In contrast, the comparison using 10 metallic coatings (two samples from each blue, brown, green, red and yellow sample group) showed that the measurement geometry could affect the spectral property of the metallic coatings. The mean colour difference between the measurements using these two instruments was ΔE^*_{ab} of 31 with a range from 22 to 41. These suggest that the measurement geometry differences did not affect the ColorChecker DC much, but did affect the metallic-coating samples. Figure 5-16 shows the colour distribution of the 106 samples and the 166 patches in the ColorChecker DC plotted in a CIELAB L^*C^* diagram and a CIELAB a^*b^* diagram.

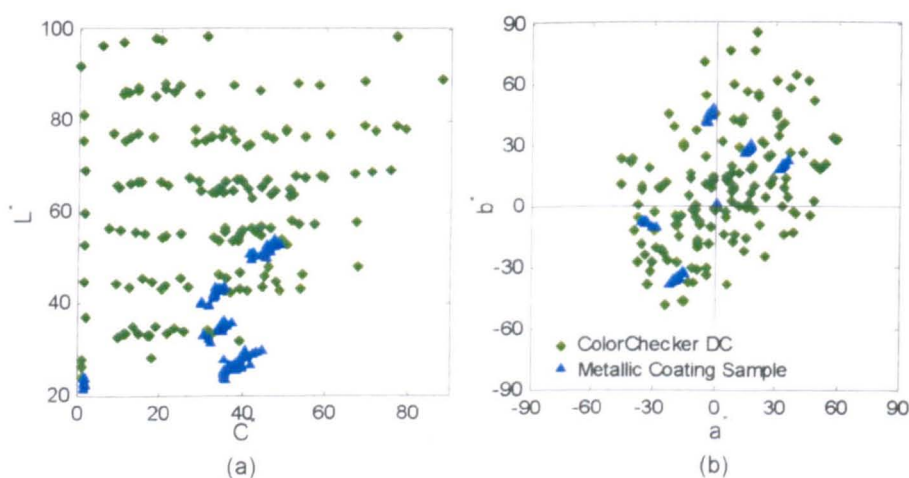


Figure 5-16: Colour distribution of the 106 metallic-coating samples and the 166 patches in the ColorChecker DC. (a) A CIELAB L^*C^* diagram. (b) A CIELAB a^*b^* diagram.

Three data sets used as training data are summarised in Table 5-18. The first training data set (Data Set 1) consisted of the mean HDR values for each of the 106 metallic-coating samples and the XYZ values of the corresponding samples. The

second data set (Data Set 2) included the mean HDR values for each of the 166 patches in the ColorChecker DC and the XYZ values of the corresponding patches. The third data (Data Set 3) were the mean HDR values for each of the 106 samples and the 166 patches and the XYZ values of the corresponding samples and patches.

The test data set consisted of the 106 metallic-coating samples. Therefore, in case of Data Set 1, a leave-one-out method was applied. The 105 metallic-coatings samples out of the 106 samples were used as training data and the remaining one sample was used to test the performance. Consequently, 106 trials needed to be made in order to evaluate a model with all test data. Similarly, for Data Set 3, all the 166 patches and the 105 samples out of the 106 samples were used as training data and the remaining one sample was used to test the performance. Again, 106 trials needed to be made in order to evaluate a model with all test data.

Table 5-18: Training and test data for characterisation.

	Data Set 1	Data Set 2	Data Set 3
Training Data	106 metallic-coating samples	166 patches in ColorChecker DC	106 metallic-coating samples + 166 patches in ColorChecker DC
Test Data	106 metallic-coating samples	106 metallic-coating samples	106 metallic-coating samples
Test Method	Leave-one-out	-	Leave-one-out

The performance of the models was evaluated in terms of CIELAB ΔE^*_{ab} between the measured and predicted values for both the training and test data sets. The results are given in Table 5-19, 5-20 and Table 5-21 for Data Set 1, 2 and 3 respectively. It can be seen that the model derived from Data Set 1 using $M = 3 \times 20$ provides best fit which had a median ΔE^*_{ab} value of 0.46. This model performed better than the other models at a significance level ($p < 0.05$) according to the Wilcoxon signed-rank test (see Section 2.10.5). As in Section 3.4.2.1, the reasons that Data Set 1 performs well are that Data Set 1 has the same surface material as that of the test data set and the colour distribution of Data Set 1 is concentrated in an area where the colours of the test data are in colour space (because both training and test data consisted of the metallic-coating samples). The very low error values in Table 5-19 demonstrate that it is possible to obtain an accurate transformation from the HDR values to the XYZ values. However, in this study, a model is needed that can predict a wider range of colours (wider than in

Data Set 1) and therefore the model derived from Data Set 1 is rejected despite its low error values in Table 5-19. In Figure 5-16, it is evident that the colour distribution of the metallic-coating samples is relatively small. However, this only considers the spatially averaged mean values; the earlier analysis demonstrated that the largest pixel value in the HDR images of the samples was about 3 to 4 times larger than even white patches. Hence, to predict such a wide range of colours, Data Set 3 (the 105 samples and the 166 patches in the ColorChecker DC) was considered to be most appropriate because the training set covered such a wide colour gamut and the test errors were smaller than those from Data Set 2. Among the models derived from Data Set 3, the model of $M = 3 \times 10$ provided the best performance of $3.39 \Delta E^*_{ab}$ (median error). However, if higher powers than needed are included, the polynomial may diverge rapidly from smooth behaviour outside the range of the training data. Since an HDR system was used in this study, extrapolation was unavoidable. Hence, the models' behaviours when the extrapolation performed need to be examined.

Figure 5-17 compares the predicted Y tristimulus values from the models of $M = 3 \times 10$ and $M = 3 \times 3$ (a linear model). It showed the comparison in terms of the median, maximum and minimum values in each image of the 106 metallic-coating samples. The median and minimum values seem to be within the interpolation process, however some of the maximum values are far beyond the range of the training data and it means that the extrapolation was applied. Although the model of $M = 3 \times 10$ was higher order than the linear model of $M = 3 \times 3$, there were only slight differences between them. However, their differences were significant ($p < 0.05$) in absolute values, although they were highly related linearly. In order to see the influence of these differences, both of the polynomial ($M = 3 \times 10$) and linear ($M = 3 \times 10$) models derive from all the 106 metallic-coating samples and the 166 patches in ColorChecker DC were applied.

After the HDR values in the images were converted into the XYZ values, only the luminance channel (namely Y tristimulus value) was processed to extract the glint information. As mentioned in Sections 2.2.2 and 2.2.3, the properties of human perception has better resolution for the achromatic channel than for the chromatic channels. Additionally, even in coloured samples, a close visual inspection reveals that the bright spots appeared to be achromatic. Certainly, it was evident that even if the spot was not perfectly achromatic, they possess a much higher luminance level than their background. Therefore, it is suggested that the luminance channel would be more important than the chromatic channels.

Table 5-19: Model performances using the Data Set 1.

	Training				Test			
Data	105 metallic-coating samples				106 metallic-coating samples			
Size of M	Training Error (ΔE^*_{ab})				Test Error (ΔE^*_{ab})			
	Median	Mean	Max	Min	Median	Mean	Max	Min
3 × 3	3.69	4.28	8.82	0.31	3.78	4.40	8.85	0.34
3 × 4	3.62	4.07	8.73	0.22	3.76	4.22	8.77	0.24
3 × 5	1.68	2.56	14.80	0.19	1.76	2.71	14.93	0.26
3 × 9	0.73	1.00	3.49	0.07	0.80	1.11	3.57	0.14
3 × 10	0.63	0.82	3.21	0.03	0.70	0.92	3.60	0.13
3 × 11	0.58	0.76	3.39	0.02	0.63	0.87	4.08	0.05
3 × 20	0.38	0.45	1.99	0.03	0.46	0.59	2.71	0.08
3 × 35	0.32	0.36	1.42	0.02	0.50	0.75	4.83	0.04
3 × 56	0.22	0.24	0.89	0.01	0.54	1.13	13.47	0.13

Table 5-20: Model performances using the Data Set 2.

	Training				Test			
Data	166 patches in the ColorChecker DC				106 metallic-coating samples			
Size of M	Training Error (ΔE^*_{ab})				Test Error (ΔE^*_{ab})			
	Median	Mean	Max	Min	Median	Mean	Max	Min
3 × 3	4.66	5.83	22.99	0.48	3.87	5.74	18.62	0.42
3 × 4	4.42	5.12	22.77	0.73	7.09	9.17	19.49	5.05
3 × 5	3.81	4.73	22.74	0.57	7.42	10.01	20.37	5.84
3 × 9	4.10	5.10	21.67	0.35	4.33	6.62	18.81	1.14
3 × 10	3.50	4.33	22.99	0.42	7.85	9.63	19.38	5.51
3 × 11	3.63	4.31	23.00	0.42	8.00	10.16	19.00	6.06
3 × 20	3.15	3.91	22.98	0.46	7.74	10.08	21.11	5.15
3 × 35	2.76	3.37	21.86	0.25	8.43	11.18	21.28	4.08
3 × 56	2.42	2.94	21.41	0.02	9.34	11.90	20.19	5.48

Table 5-21: Model performances using the Data Set 3.

	Training				Test			
Data	166 patches in the ColorChecker DC + 105 metallic-coating samples				106 metallic-coating samples			
Size of M	Training Error (ΔE^*_{ab})				Test Error (ΔE^*_{ab})			
	Median	Mean	Max	Min	Median	Mean	Max	Min
3 × 3	4.43	5.76	22.99	0.13	3.83	5.55	17.75	0.58
3 × 4	4.53	5.90	21.19	0.45	4.49	6.38	16.84	2.23
3 × 5	4.06	5.49	21.13	0.25	3.79	5.72	16.85	1.73
3 × 9	3.93	5.50	25.02	0.37	3.54	5.71	15.88	1.03
3 × 10	3.82	5.31	28.16	0.51	3.39	5.29	15.55	1.43
3 × 11	4.17	5.36	30.61	0.50	4.29	5.45	14.87	1.49
3 × 20	4.06	5.20	22.37	0.24	4.18	5.43	15.05	1.96
3 × 35	3.68	4.77	20.17	0.35	4.07	5.56	12.94	1.51
3 × 56	2.81	3.86	20.62	0.04	3.98	4.37	9.74	0.44

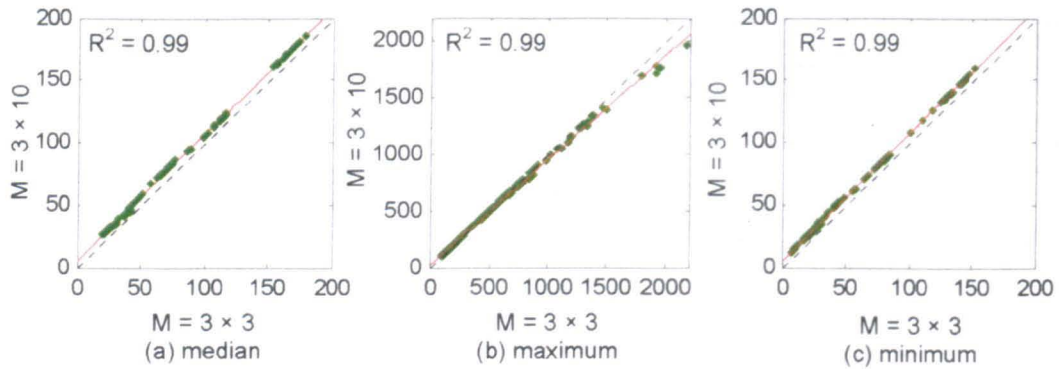


Figure 5-17: Comparisons of the predicted Y values from the models of $M = 3 \times 10$ and $M = 3 \times 3$. (a) Median values for each of the 106 samples. (b) Maximum for each of the 106 samples. (c) Minimum for each of the 106 samples.

5.3.4 Glint Feature Extraction

The observers were instructed to scale the glint of the metallic-coating samples according to both the observed contrast between bright spots (sparkles) and their background (surround), and the amount (number) of spots (see Section 5.2.3). It is clear that there are two sub-parameters; local contrast and the amount of spots, although the relative importance that observers gave to these sub-parameters is unknown. It suggests that the image should be analysed in terms of contrast rather than absolute intensity of the pixels in the image. Based on contrast analysis, the pixels corresponding to spots were identified and then segmented. Finally, a feature corresponding to glint was extracted based on the premise that the number of such segmented pixels or agglomerates of such pixels is proportional to the perceptual glint.

5.3.4.1 Contrast Measure

Figure 5-18 shows the range of maximum and minimum luminance values in each image together with its mode (the value most frequently occurs in an image). In Figure 5-18 each vertical line represents the luminance range of one metallic-coating sample; the horizontal black dashed line indicates the luminance value of the white patch for reference. It can be seen that, for example, the minimum values of some of the yellow colour samples are much higher than that of other samples and sometimes even higher than the maximum values of some of the samples. In spite of the low maximum values of some of the samples, the scale values of the perceptual glint for these samples were

not one (in visual assessment, one indicated non-glint, see Section 5.2.3). Meanwhile, the spots and the background were both observed from those yellow samples, whose minimum values were higher than the maximum values of some of the samples. Thus, it is evident that the spots cannot be identified in terms of their absolute luminance values; rather, it is luminance contrast that is important. Therefore, the pixel values were represented in terms of the intensity of contrast by subtracting the mode of each image from every pixel value in the image. It is more common to subtract a mean of an image rather than a mode in order to measure contrast. However, the contrast which should be measured here is the difference between the pixel values and the background (no spots area); the mode is considered to be more representative than the mean for the background intensity (the further detail is explained in Section 5.3.4.2.2).

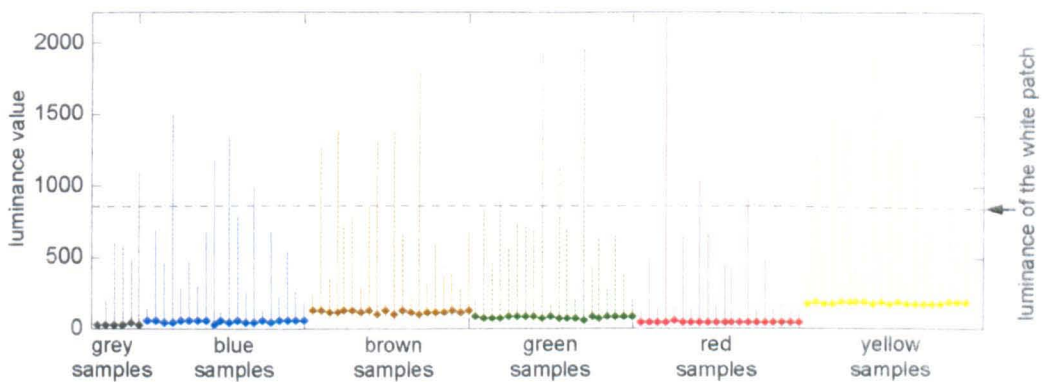


Figure 5-18: Range of maximum and minimum luminance values in each image together with the mode (filled circle) – the luminance values were obtained from the linear characterisation model. The horizontal black dashed line indicates the luminance value of the white patch for reference.

5.3.4.2 Segmentation

Image segmentation was implemented in order to separate the image into two regions: spots and background. The spots on the metallic-coating sample are the bright regions which are caused by the characteristics of the aluminium flakes contained in the coating. However, it should be noted that the spots do not have the actual physical size of the reflecting flakes (see Chapter 4). The background is defined as the region where there are no spots.

A *global thresholding technique* partitions an image histogram by using a single threshold t as illustrated in Figure 5-19. This segmentation technique is effective, if an image contains objects and background, in such a way that objects and background pixel values have levels grouped into two dominated and well-separated modes.

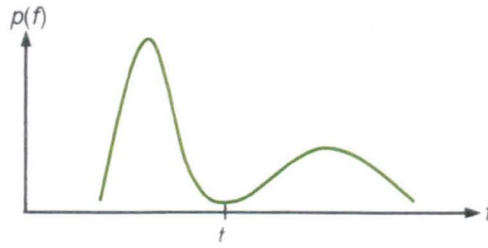


Figure 5-19: A histogram of an image that can be partitioned by a single threshold t .

A simple way to separate objects from background is to select a single threshold t that separate the modes of the two distributions. If $I(i, j)$ is the pixel value at position (i, j) , where i ($i = 1, 2, \dots, m$) and j ($j = 1, 2, \dots, n$) represent the ranges of position (m and n are the number of rows and columns in an image), then any pixel in an image where $I(i, j)$ is larger than (or equal to) the threshold is classed as an object; otherwise the pixel is classed as background. The thresholded image, $g(i, j)$ is often defined using a binary indication as given in Equation 5-7 (Gonzalez & Woods, 1992).

$$g(i, j) = \begin{cases} 1 & \text{if } I(i, j) \geq t \\ 0 & \text{if } I(i, j) < t \end{cases} \quad \text{Equation 5-7}$$

For this simple point-wise binary operation, the choice of threshold needs to be made. For the case where an image histogram contains clearly distinguished modes, it is relatively straightforward to define an optimal (or effective) threshold value. However, if the two distributions are not clearly separated, it is more difficult to define an optimal threshold. In Figure 5-20, the luminance histograms for six images, each corresponding to a metallic-coating sample from one of the colours, are given. It can be seen that each distribution contains only one mode. The mode is generally located towards the lower end of the luminance range in each case so that it is assumed that the mode represents pixels belonging to the background. The intensity of the spots are widely scattered and it is difficult to differentiate spot pixel values from background pixel values. Note that if the background and spot distributions overlap, as seems likely in Figure 5-20, then a perfect segmentation by thresholding is impossible. However, an optimal threshold value is sought. Therefore, in order to find an optimal threshold, two approaches were considered: an *iterative method* (Gonzalez and Woods *et al.*, 2004) and a *distribution-estimate method*. The criteria used to evaluate the effectiveness of the segmentation methods were the correlation measures, R , R^2 and RMSE between the scale values of the perceptual glint (see Section 5.2.4.1) and the outputs of the glint model (that

incorporates the thresholding). This model is described in Sections 5.3.4.3 and 5.3.4.4. However, the two threshold-estimation methods are first briefly described.

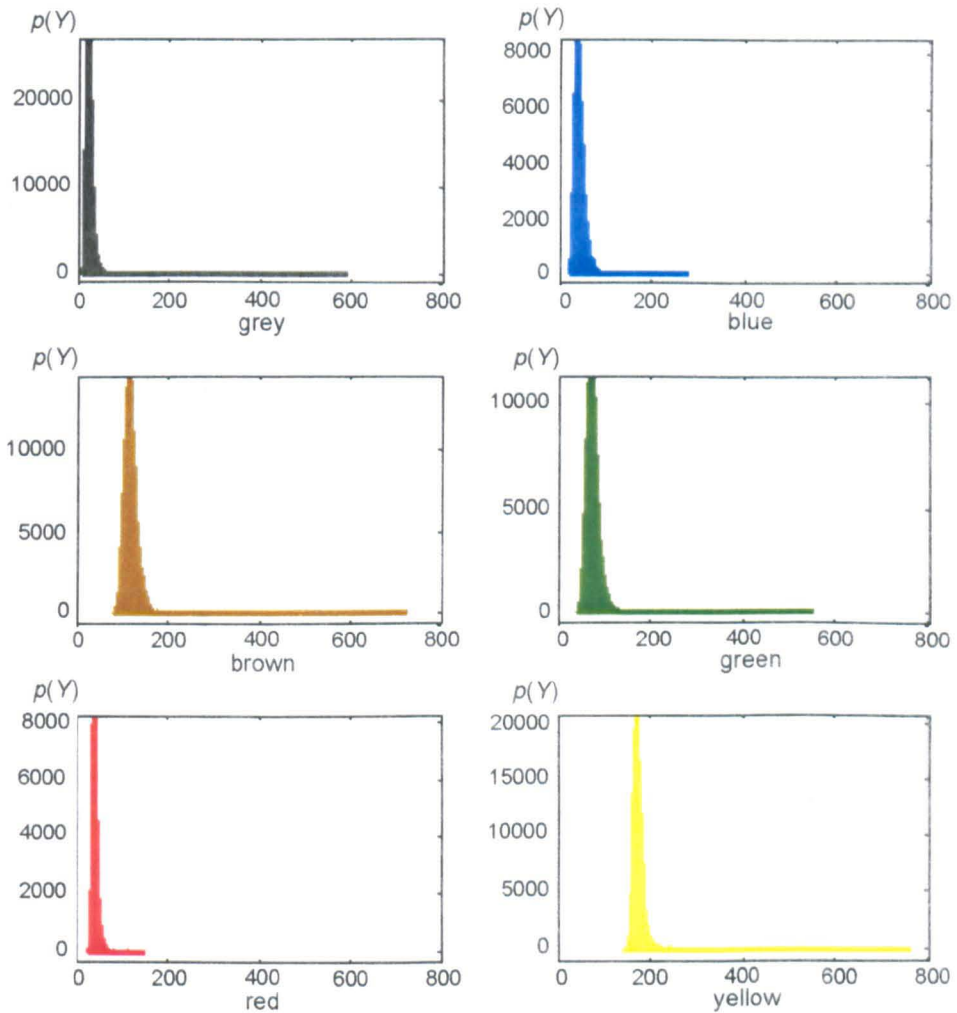


Figure 5-20: Luminance channel's histograms of one of the images from each grey, blue, brown, green, red and yellow colour samples (the luminance value was obtained using the linear characterisation model).

5.3.4.2.1 Iterative Method

The *iterative method* (Gonzalez and Woods *et al.*, 2004) chooses a threshold t_i by exhaustive search, picking different thresholds until one is found that produces a satisfactory result as judged by certain criteria. In this study, the criteria were the correlation measures described in the previous section.

5.3.4.2 Distribution-Estimate Method

This *distribution-estimate method* is based on the assumption that the distributions of the background of the samples should be a bell-shaped, symmetric histogram with most of the frequency counts bunched in the middle and with the counts dying off out in the tails. Histograms of solid-colour-coating panels were measured and can be seen to demonstrate this property; two examples are given in Figure 5-21. It is reasonable to assume that the histograms of the solid-colour-coating panels should be similar to that of the metallic-coating samples used in this experiment if there were no the aluminium flakes. Therefore, a threshold t_d can be determined by estimating an upper limit of the distribution of the background. As is illustrated in Figure 5-22, when p represents the mode and l is the minimum value (lower limit) in an image, t_d is estimated to be $t_d = (2p - l)$.

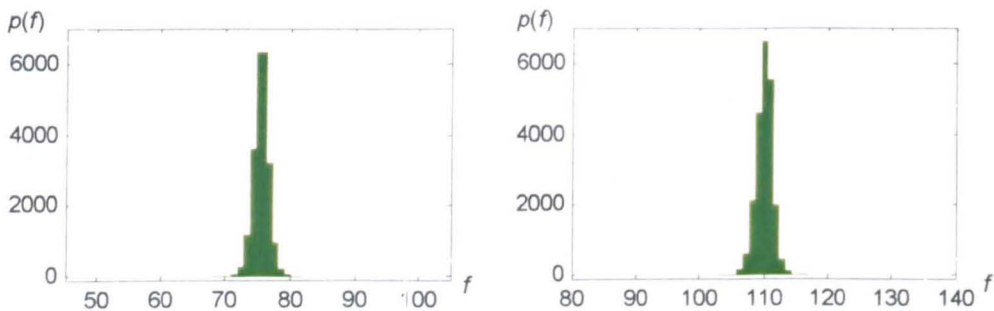


Figure 5-21: Histograms of solid-colour-coating panels.

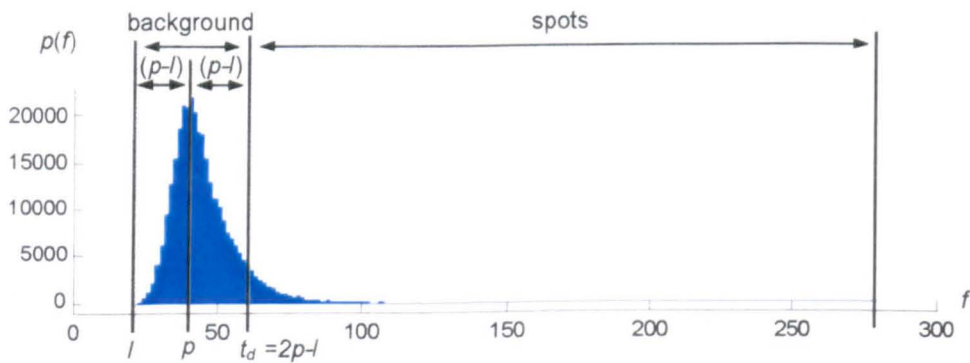


Figure 5-22: An illustration of the theory of the distribution-estimate method.

5.3.4.3 Glint Features – Statistical Approaches

Various statistical approaches were proposed in this section to extract glint features correlated with the perceptual glint based on the segmented images.

An obvious approach would be to use the number of pixels that exceed a certain threshold t as a direct measure of perceptual glint. This method is based on the implicit

assumption that perceptual glint depends only on the number of pixels that have been identified as spots and ignores the spatial distributions of these pixels. It can be informally described in Equation 5-8 using the indications; $I(i, j)$ is the luminance value of the image I at the pixel position (i, j) , p is the mode and t is either t_i or t_d . The mode was subtracted from every pixel in the image to derive the contrast measure (see Section 5.3.4.1). This metric is called $M1$.

$$M1 = \text{count}((I(i, j) - p) \geq t) \quad \text{Equation 5-8}$$

However, although the metric $M1$ is simple, it does not differentiate between pixels that are just over the threshold and those that greatly exceeded the threshold. Therefore, the metric $M2$ computed the sum of those pixel values that belonged to the spots, as given in Equation 5-9.

$$M2 = \sum((I(i, j) - p) \geq t) \quad \text{Equation 5-9}$$

In the metrics $M1$ and $M2$, individual pixels were counted as one element of the spots, although more than one pixel might compose a particle as one spot in the image. In fact, the physical size of the aluminium flakes has a diameter of between 5 and 50 μm which is much smaller than the physical size of a pixel ($114.5 \times 114.5 \mu\text{m}$) in this experimental setting. So, one spot may consist of one pixel. However, the apparent size of one spot may exceed the actual physical size of reflecting flakes (Chapter 4). It was also mentioned by Đuriković (2003) that the bright sparkles observed on metallic coatings looked much larger than the physical size of flakes. Therefore, instead of the number of pixels, the metric $M3$ counted particles which consisted of one or more pixels above threshold in the image. Particles were identified by labelling 8-connected components (Gonzalez and Woods *et al.*, 2004). By scanning the image pixel-by-pixel, if a pixel value at the position (i, j) is less than the threshold, simply move on to next scanning position. If the value is over than the threshold, examine 8-neighboring pixels. If none of the neighbouring pixels is over threshold, the pixel $I(i, j)$ is identified as a particle consisting of only one pixel. However, if any of the 8-neighbors are over the threshold, those pixels and $I(i, j)$ are labelled as elements that compose a particle. Figure 5-23 (a) indicates the positions of the 8-neighboring pixels of a pixel at the position (i, j) and Figure 5-23 (b) is an example of a labelled image. Three connected components (*i.e.*, three particles in this study) can be found in the image. Consequently, the metric $M3$ is given in Equation 5-10.

$$M3 = \text{count}(\text{particle}((I(i, j) - p) \geq t)) \quad \text{Equation 5-10}$$



Figure 5-23: (a) The position of 8-neighboring pixels of a pixel at coordinates (i, j) (8 shadowed cells). (b) An Example of a labelled image. It indicates three connected components in an image.

The metric $M4$ computed the sum of mean value of each particle as given in Equation 5-11.

$$M4 = \sum \text{mean}(\text{particle}((I(i, j) - p) \geq t)) \quad \text{Equation 5-11}$$

The metrics $M1$, $M2$, $M3$ and $M4$ introduced above already take into account the contrast in the image by subtracting the mode of the image. However, in addition to this, the metrics $M5$, $M6$, $M7$ and $M8$ adopted another contrast effect which is described by Weber's law (see Section 2.2.3.1). The metrics $M5$, $M6$, $M7$ and $M8$ are corresponding to the metrics $M1$, $M2$, $M3$ and $M4$ as given in Equation 5-12.

$$M5 = \frac{M1}{p} ; M6 = \frac{M2}{p} ; M7 = \frac{M3}{p} ; M8 = \frac{M4}{p} \quad \text{Equation 5-12}$$

5.3.4.4 Non-Linearity

The correlation of the outputs of the metrics $M1-M8$ with perceptual glint should also be considered. The outputs of the metrics may not necessarily correlate with perceptual glint linearly. However, final model predictions were expected to correlate linearly with the perceptual glint. Therefore, a non-linearity stage was incorporated to find a linear approximation function of the metrics $M1-M8$ to the scale values of the perceptual glint (see Section 5.2.4.1). Applied forms were linear, power, exponential and logarithm transformation as shown in Equation 5-13.

<i>Linear</i>	$SV = \alpha M + \beta$	
<i>Power 1</i>	$SV = \alpha M^\gamma$	
<i>Power 2</i>	$SV = \alpha M^\gamma + \beta$	Equation 5-13
<i>Exponential</i>	$SV = \alpha \times \exp(\gamma M) + \beta$	
<i>Logarithm 1</i>	$SV = \alpha \times \log_{10}(M) + \beta$	
<i>Logarithm 2</i>	$SV = \alpha \times 10^M + \beta$	

where SV indicates the scale value of the perceptual glint, M indicates the output of the metrics and α , β and γ are parameters.

5.4 Model Performance

Performance of the models from all combinations of the segmentation methods, the metrics $M1-M8$ and the linearisation functions was evaluated by measuring a linear correlation with the scale values of the perceptual glint. R , R^2 and RMSE between the model predictions and the scale values of the samples were utilised as measures of the performance of the models. Prior to computing these error measures, both the model predictions and the scale values were normalised in the range 0-1 so as to compare RMSE measures from the various models with the same unit.

A flowchart in Figure 5-24 summarises the framework of the proposed models and the evaluation procedure of the model performance.

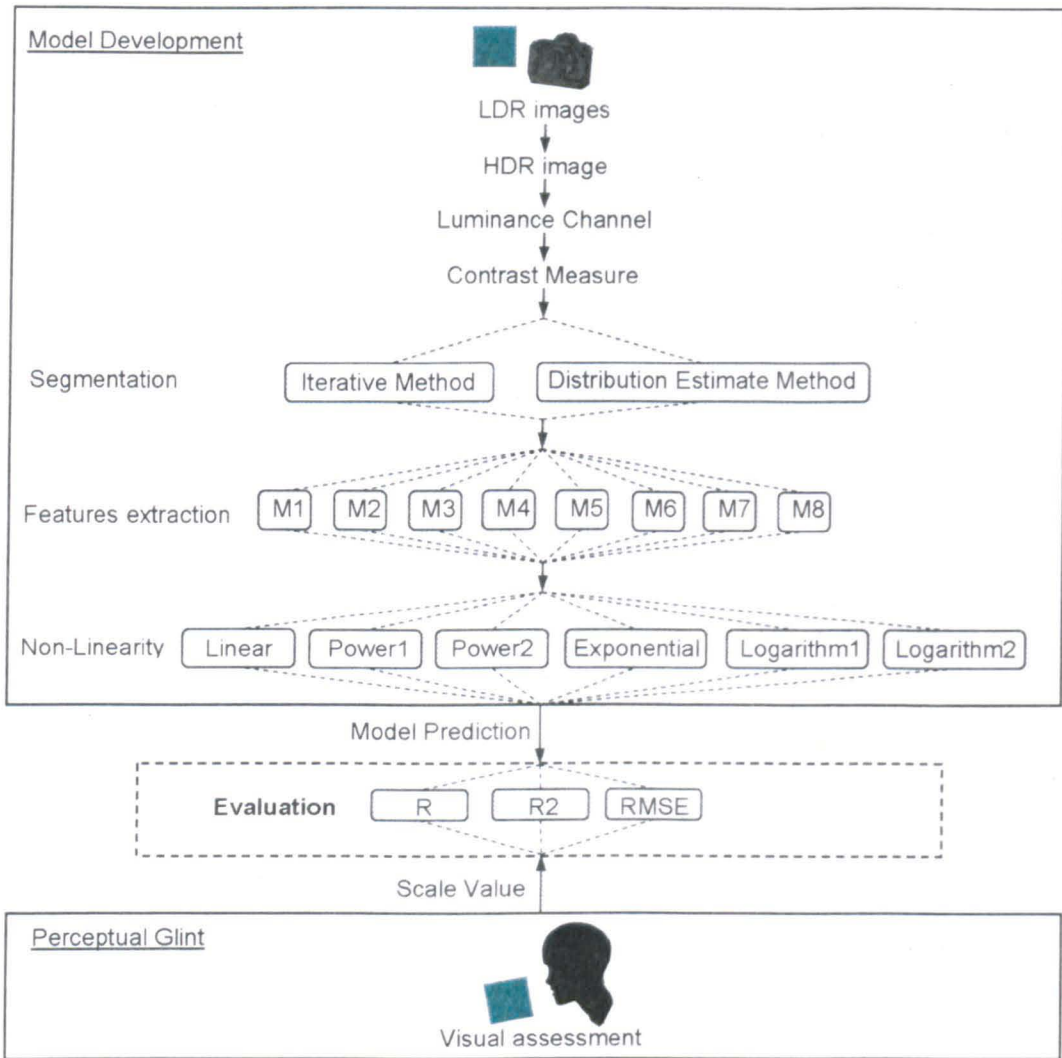


Figure 5-24: A flowchart of a summary of the model components and their evaluation processes.

5.4.1 Model Performance: Iterative Method

The models were implemented using the iterative segmentation method (see Section 5.3.4.2.1). A sequence of threshold t_i values was applied to all the samples at each trial. Thresholds varied from zero upwards until the threshold exceeded the intensities of all pixels in one of the images in the set. An example is given in Figure 5-25, the iteration is terminated at the threshold t_{ik} , when all the pixels of a red colour sample become below the threshold. Note that this procedure applied to the images after their mode were subtracted from the every pixel of each image so that the pixel value of zero indicates the mode; in other words, the mean value of the background.

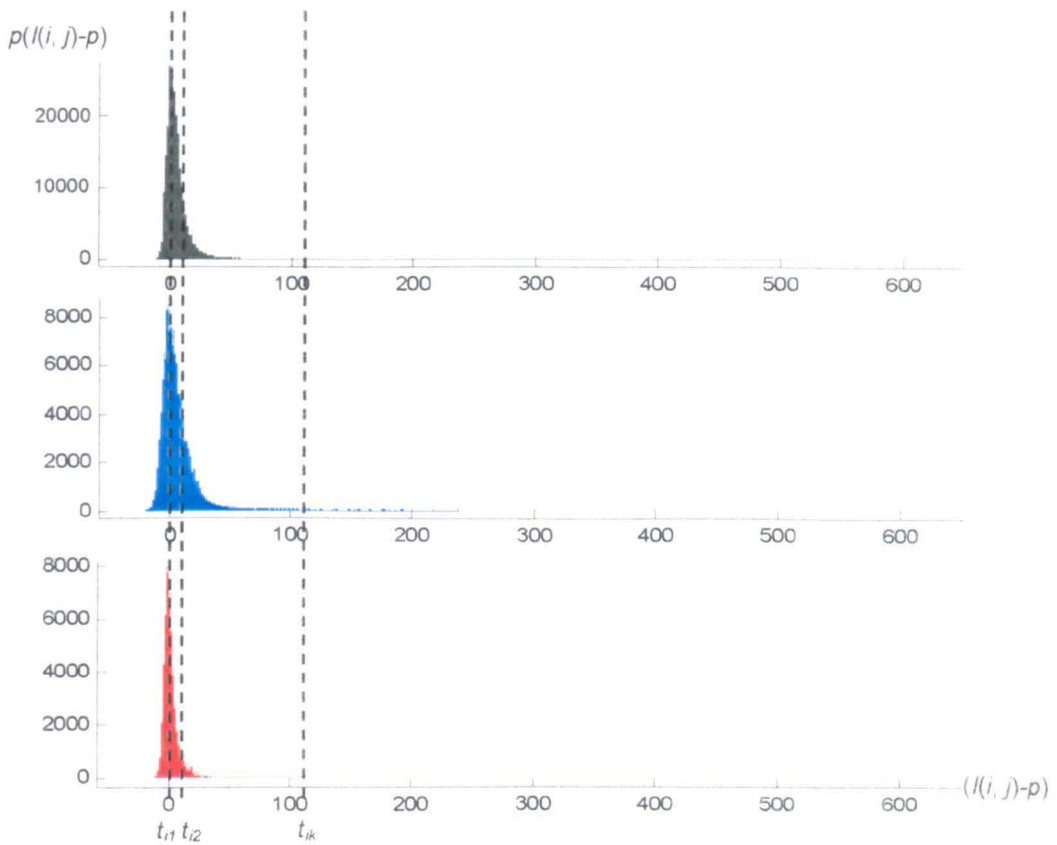


Figure 5-25: A threshold t_i with the image histograms of three samples (grey, blue, red colour samples). The iteration is terminated at the threshold t_k , when all the pixels of the red colour sample become below the threshold.

Figure 5-26 and Figure 5-27 show the prediction errors of the models associated with the metrics $M1-M8$ plotted at each threshold. A series of six graphs on the left column shows the errors in terms of R ; the graphs in the middle present the errors in terms of R^2 ; and the right graphs are for RMSE. A set of three graphs in each row presents the each of the six linearisation functions utilised. Figure 5-26 and Figure 5-27 present the prediction errors when the models applied to samples whose luminance values were transformed from the HDR values using the linear regression characterisation model and the polynomial regression characterisation model respectively. Top six models in terms of their performance in R^2 and RMSE are given in Table 5-22 based on the samples associated with the linear model and Table 5-23 based on the samples associated with the polynomial models.

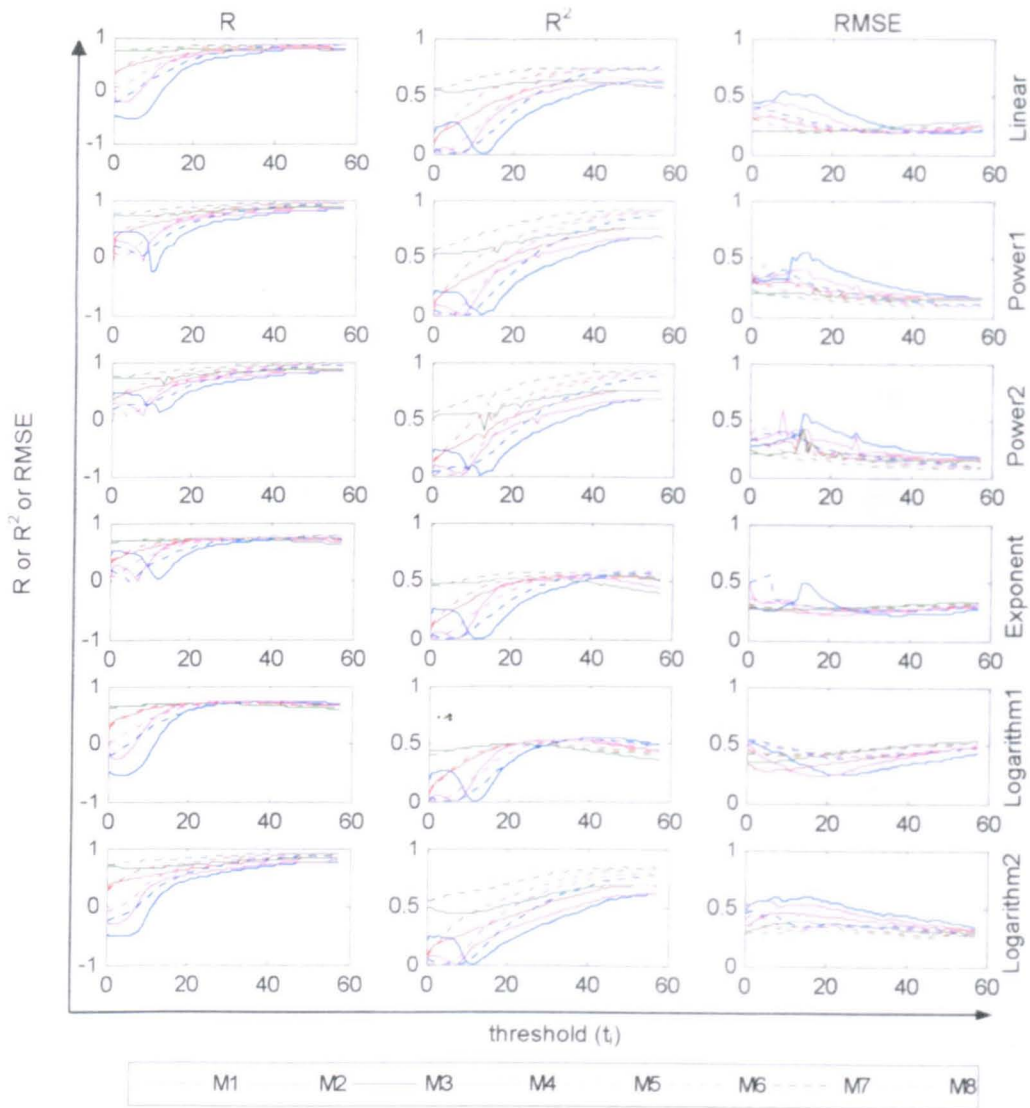


Figure 5-26: The prediction errors of the models associated with the metrics M1-M8 at each threshold. A series of six graphs on the left, middle and right columns shows the errors in terms of R , R^2 and RMSE respectively. A series of three graphs in each row indicates the each of the six linearisation functions. The vertical axis presents the error and the horizontal axis indicates threshold t value. The results are based on the luminance values transformed from the HDR values using the linear regression characterisation model.

Table 5-22: Top six models in terms of their performance in R^2 and RMSE. The results are based on the luminance values transformed from the HDR values using the linear regression characterisation model.

R^2					RMSE				
Rank	Metric	Nonlinearity	R^2	t_i	Rank	Metric	Nonlinearity	RMSE	t_i
1	M6	Power2	0.94	56	1	M6	Power2	0.078	56
2	M6	Power1	0.93	56	2	M5	Power2	0.087	56
3	M5	Power2	0.92	56	3	M8	Power2	0.090	56
4	M5	Power1	0.91	57	4	M6	Power1	0.092	57
5	M8	Power2	0.91	57	5	M5	Power1	0.092	57
6	M8	Power1	0.90	57	6	M8	Power1	0.078	56

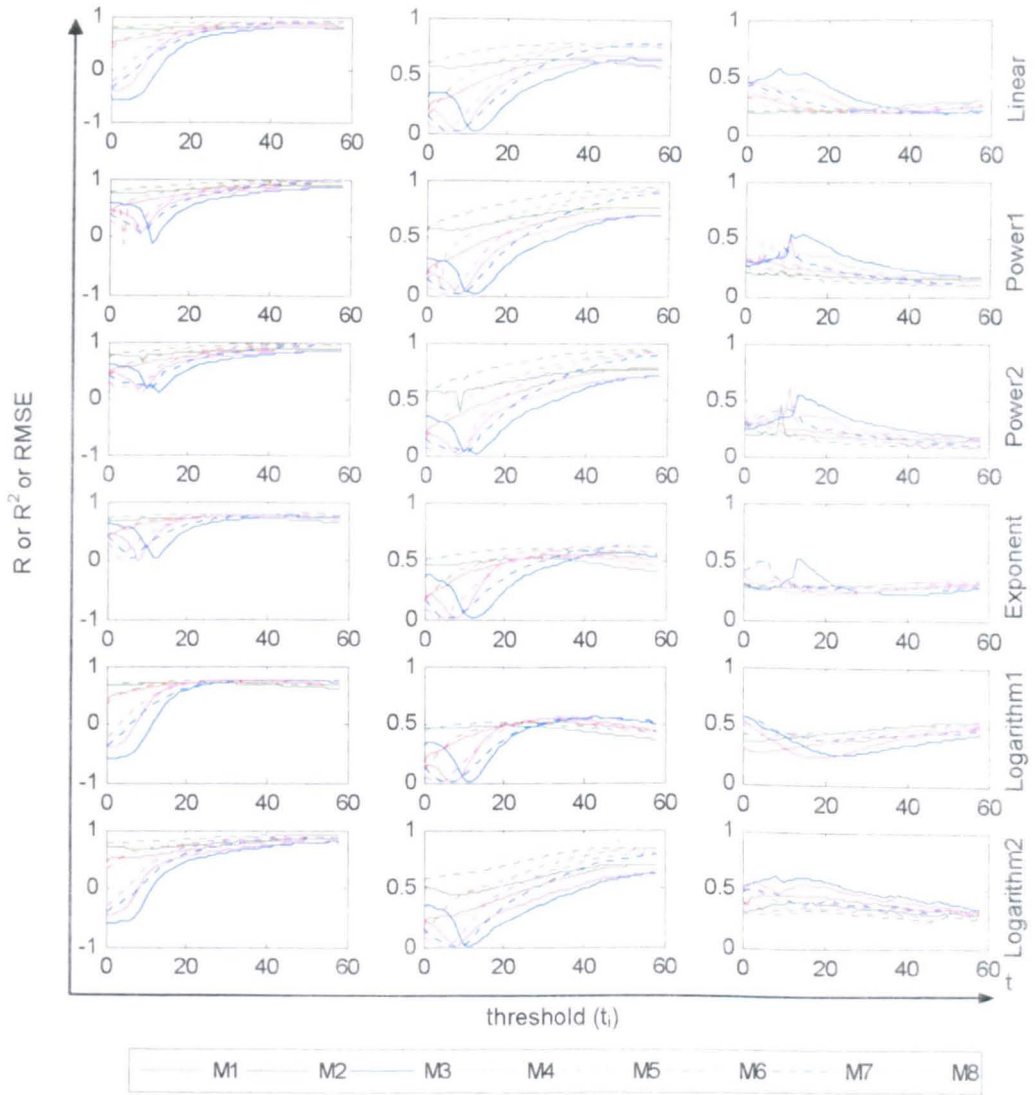


Figure 5-27: The prediction errors of the models associated with the metrics M1-M8 at each threshold as well as Figure 5-26. The results are based on the luminance values transformed from the HDR values using the polynomial regression characterisation model.

Table 5-23: Top six models in terms of their performance in R^2 and RMSE. The results are based on the luminance values transformed from the HDR values using the polynomial regression characterisation model.

		R^2			RMSE				
Rank	Metric	Nonlinearity	R^2	t_i	Rank	Metric	Nonlinearity	RMSE	t_i
1	M6	Power2	0.94	57	1	M6	Power2	0.080	58
2	M6	Power1	0.93	58	2	M5	Power2	0.083	57
3	M5	Power2	0.93	58	3	M8	Power2	0.088	57
4	M5	Power1	0.92	58	4	M6	Power1	0.093	58
5	M8	Power2	0.91	58	5	M5	Power1	0.095	58
6	M8	Power1	0.91	58	6	M8	Power1	0.095	58

It can be seen from Figure 5-26, Figure 5-27, Table 5-22 and Table 5-23 that the model with the metric *M6* and the linearisation using *Power2* provides the best fit to the perceptual glint. The R^2 value of 0.94 and the RMSE value of 0.078 were obtained for this model based on the samples associated with the linear characterisation model, and the R^2 value of 0.94 and the RMSE value of 0.080 were found based on the samples associated with the polynomial characterisation model. The predictions of these models for individual samples were given in Appendix IV. There was no significant difference in the model predictions between the results based on the linear characterisation models and those based on the polynomial characterisation models ($p>0.05$). The model with *M5* and *Power2* also showed the excellent prediction in terms of both R^2 and RMSE values. This case also did not show any differences caused by the characterisation models ($p>0.05$). These results indicate that the choice of characterisation model does not affect the performance of the glint models. Note that although there was a significant difference between the linear and polynomial models in their performance (see Section 5.3.3), their predictions were highly correlated linearly (Figure 5-17).

These encouraging results from *M5* and *M6* give evidence that it is sufficient to analyse the individual pixels in the image and not necessary to consider the particle which was more complicated. The comparisons of the results of *M5* with *M1* and that of *M6* with *M2* reveal the necessity of accomplishing normalisation accordance with Weber's law. Figure 5-28 presents the comparisons of the perceptual glint with the model predictions of *M1* and *M2* (before Weber's law applied), and *M5* and *M6* (after Weber's law applied) at the threshold $t_i = 56$. These model predictions were based on the luminance values of the samples obtained from the linear characterisation model. The model predictions for *M1*, *M2*, *M5* and *M6* given in Figure 5-28 (a, b, c and d) are the model outputs before the linearisation is applied, but in Figure 5-28 (e) and (f) for *M5* and *M6* are the results of the linearisation using *Power2*. Also, both the scale values of the perceptual glint and the model predictions are not normalised in the range 0-1. Figure 5-28 (a) and (b) are evident that there are linear correlations between the perceptual glint and the model predictions with *M1* and *M2* for the samples having a same colour, however, it shows a wide disparity between the colour groups. Compared with that, the performance of *M5* and *M6* (Figure 5-28 (c, d, e and f)) reveals that a normalisation using the mode of the image is essential. This is because of the fact that although two samples had similar perceptual glint, the glint predictions from *M1* and *M2* were much higher for the sample whose mean background was brighter than the

other. This is explained by the human visual system which is Weber's law and this was accomplished by the normalisation.

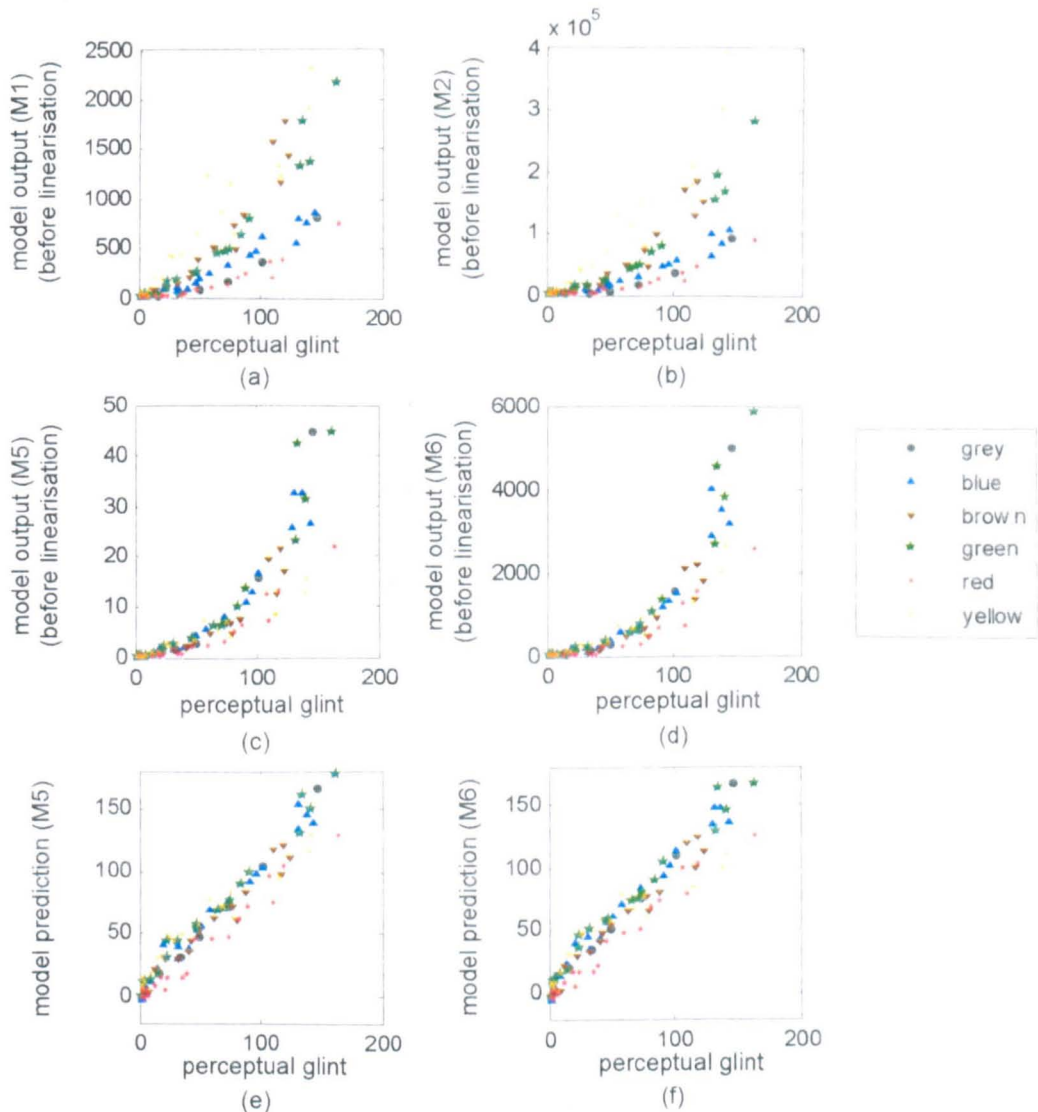


Figure 5-28: Comparisons of the perceptual glint against with the model predictions at the threshold of 56 (based on the luminance values obtained from the linear characterisation model). (a, b, c and d) The predictions of the model with of M1, M2, M5 and M6 respectively (before linearisation). (e and f) The predictions of the model with of M5 and M6 respectively (after the linearisation using Power2).

Although this model incorporated with M6 could predict the perceptual glint very well, there is no clear indication of the optimal threshold. In other words, in Figure 5-26 and Figure 5-27 where the model performance are presented as a function of threshold, there is no clear peak of R and R^2 nor valley for RMSE to indicate an optimal threshold. Although such model predicted the perceptual glint well at the wide range of the threshold values, these thresholds might be limited to the set of the samples used in this experiment. For example, the model performance at the wide range of the threshold

around 40–58 had considerably linear correlation with the perceptual glint in this experiment, whereby these threshold can be applied to any sample whose colour and glint level are similar to the samples used in this experiment. However, there is no guarantee for performance when it is applied to samples whose colour and glint level are very different from the samples used in this experiment. Therefore, this method incorporated with the iterative segmentation method is considered to be sample-set dependent.

The final model with *M6* and *Power2* incorporated with the iterative method is summarised in Equation 5-14.

$$Glint\ Model = \alpha \times \left(\frac{\sum(I(i, j) - p) \geq t_i}{p \times m \times n} \right)^\gamma + \beta \quad \text{Equation 5-14}$$

where $I(i, j)$ is a pixel value (luminance value) of an image I at the pixel position of (i, j) ; $(i=1,2,\dots,m)$ and $(j=1,2,\dots,n)$, n and m represent the numbers of row and column in a image, p is the mode in an image, t_i is the threshold and α , γ and β are model parameters.

5.4.1.1 Model Performance: Distribution-Estimate Method

The models were implemented by incorporating with the distribution-estimate segmentation method (see Section 5.3.4.2.2). This segmentation method identifies a threshold of individual samples by estimating the distribution of the background. It assumes that the range of the distribution of pixel values belonging to the background is twice a range between the mode and the minimum value in the image. In this method, the minimum value in the image can be used an indicator of one end of the distribution. However, it might not belong to the background distribution. In practice, an image often contains noise, especially at both ends of a dynamic range of imaging systems. Therefore, when in doubt, it is necessary to precede noise reduction by applying such as spatial filtering: median, lowpass filtering, *etc.*, before determining the minimum value of the background distribution. However, an examination of the images of the samples led to the conclusion that noise was relatively unimportant in the images of the samples used in this experiment. Figure 5-29 shows the minimum values in each image plotted against their mode. It can be seen that the minimum values increase with the mode. A possible reason for the low noise level in this system may be that some noise is already removed during the spatial correction applied during creating the HDR images. Also, because HDR images are being used, around the minimum values in the HDR image

samples are captured relatively in the middle range of the LDR imaging system where is always less noisy than both end of the dynamic range of the system. Consequently, no noise removal process was applied to the images.

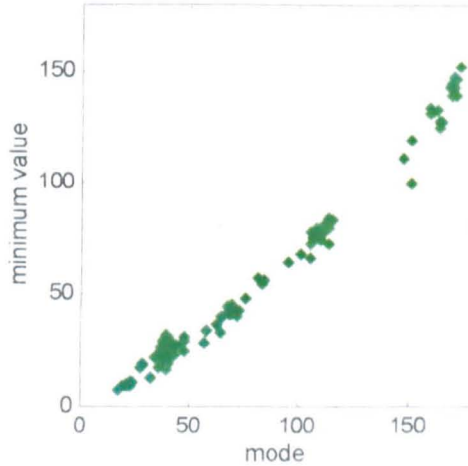


Figure 5-29: The minimum values in the images of the samples were plotted against the value at their mode (based on the luminance value obtained from the linear characterisation model).

The performance of the models incorporated with each metric $M1-M8$ and the linearisation functions was given in terms of R , R^2 and RMS as summarised in Table 5-24 to Table 5-29. Table 5-24 to Table 5-26 present the results of the models applied to the samples associated with the linear characterisation model and Table 5-27 to Table 5-29 present those associated with the polynomial model. The results indicate that it is sufficient to analyse the individual pixels and not necessary to consider the particle, as well as the results obtained from the models using the iterative method.

A best model performance was the R , R^2 and RMS values of 0.92, 0.84 and 0.132 respectively, from the model associated with the metric $M2$ and $Power2$ as the linear characterisation model applied. Figure 5-30 (a) shows the predictions of this model plotted against the scale values of the perceptual glint for all the samples. The model predictions of all the samples were given in Appendix IV.

As results of applying the models to the samples associated with the polynomial characterisation mode, the model incorporated with the metric $M2$ and $Power1$ were found to be the best and the R , R^2 and RMS values of 0.92, 0.85 and 0.116 were obtained respectively. Figure 5-30 (b) plots the model predictions against the scale values. The model predictions for all the samples were given in Appendix IV.

It can be seen from Figure 5-30 (a) and (b) that the correlations weaken when the value of the perceptual glint and the model prediction increases. Although the performance at the higher glint values were less perfect, overall results of these models are satisfactory. It was also found that there were no significant differences ($p>0.05$) between the linear characterisation model and the polynomial characterisation model in terms of the predictions of these models.

Table 5-24: R values between the perceptual glint and the model predictions based on the luminance value obtained from the linear characterisation model.

R	M1	M2	M3	M4	M5	M6	M7	M8
Linear	0.79	0.91	0.62	0.43	0.63	0.78	0.49	0.71
Power1	0.80	0.92	0.64	0.44	0.65	0.84	0.48	0.71
Power2	0.80	0.92	0.64	0.43	0.66	0.84	0.49	0.70
Exponential	0.71	0.81	0.65	0.40	0.57	0.68	0.47	0.70
Logarithm 1	0.71	0.77	0.70	0.52	0.53	0.63	0.48	0.73
Logarithm2	0.72	0.88	0.54	0.48	0.60	0.81	0.43	0.63

Table 5-25: R² values between the perceptual glint and the model predictions based on the luminance value obtained from the linear characterisation model.

R ²	M1	M2	M3	M4	M5	M6	M7	M8
Linear	0.63	0.82	0.34	0.18	0.40	0.61	0.24	0.50
Power1	0.64	0.84	0.41	0.19	0.43	0.71	0.23	0.50
Power2	0.64	0.84	0.41	0.18	0.43	0.71	0.24	0.50
Exponential	0.50	0.65	0.42	0.16	0.33	0.47	0.22	0.49
Logarithm 1	0.52	0.60	0.49	0.27	0.28	0.40	0.23	0.54
Logarithm2	0.53	0.77	0.29	0.23	0.35	0.66	0.19	0.39

Table 5-26: RMSE values between the perceptual glint and the model predictions based on the luminance value obtained from the linear characterisation model.

RMSE	M1	M2	M3	M4	M5	M6	M7	M8
Linear	0.207	0.172	0.340	0.304	0.284	0.244	0.277	0.235
Power1	0.183	0.136	0.297	0.307	0.223	0.160	0.290	0.227
Power2	0.183	0.132	0.298	0.304	0.228	0.156	0.272	0.246
Exponential	0.300	0.270	0.239	0.302	0.331	0.311	0.307	0.219
Logarithm 1	0.438	0.432	0.227	0.357	0.523	0.501	0.452	0.271
Logarithm2	0.330	0.272	0.495	0.377	0.344	0.303	0.404	0.431

Table 5-27: R values between the perceptual glint and the model predictions based on the luminance value obtained from the polynomial characterisation model.

R	M1	M2	M3	M4	M5	M6	M7	M8
Linear	0.80	0.91	0.63	0.45	0.66	0.80	0.51	0.70
Power1	0.80	0.92	0.63	0.46	0.67	0.85	0.50	0.70
Power2	0.79	0.92	0.60	0.46	0.67	0.85	0.51	0.68
Exponential	0.75	0.83	0.64	0.42	0.62	0.73	0.50	0.72
Logarithm 1	0.71	0.78	0.70	0.54	0.56	0.66	0.51	0.73
Logarithm2	0.71	0.87	0.53	0.50	0.60	0.81	0.45	0.61

Table 5-28: R² values between the perceptual glint and the model predictions based on the luminance value obtained from the polynomial characterisation model.

R ²	M1	M2	M3	M4	M5	M6	M7	M8
Linear	0.64	0.82	0.39	0.20	0.43	0.65	0.26	0.49
Power1	0.65	0.85	0.39	0.21	0.45	0.73	0.25	0.49
Power2	0.63	0.84	0.37	0.21	0.46	0.73	0.26	0.46
Exponential	0.56	0.70	0.41	0.17	0.37	0.51	0.25	0.50
Logarithm 1	0.51	0.60	0.48	0.29	0.31	0.43	0.26	0.54
Logarithm2	0.51	0.75	0.28	0.25	0.36	0.66	0.20	0.37

Table 5-29: RMSE values between the perceptual glint and the model predictions based on the luminance value obtained from the polynomial characterisation model.

RMSE	M1	M2	M3	M4	M5	M6	M7	M8
Linear	0.195	0.159	0.323	0.286	0.270	0.232	0.269	0.269
Power1	0.173	0.116	0.322	0.295	0.219	0.154	0.301	0.272
Power2	0.189	0.145	0.371	0.291	0.221	0.166	0.270	0.338
Exponential	0.285	0.257	0.235	0.287	0.319	0.306	0.288	0.213
Logarithm 1	0.425	0.421	0.232	0.348	0.507	0.491	0.416	0.243
Logarithm2	0.347	0.293	0.493	0.395	0.353	0.311	0.427	0.465

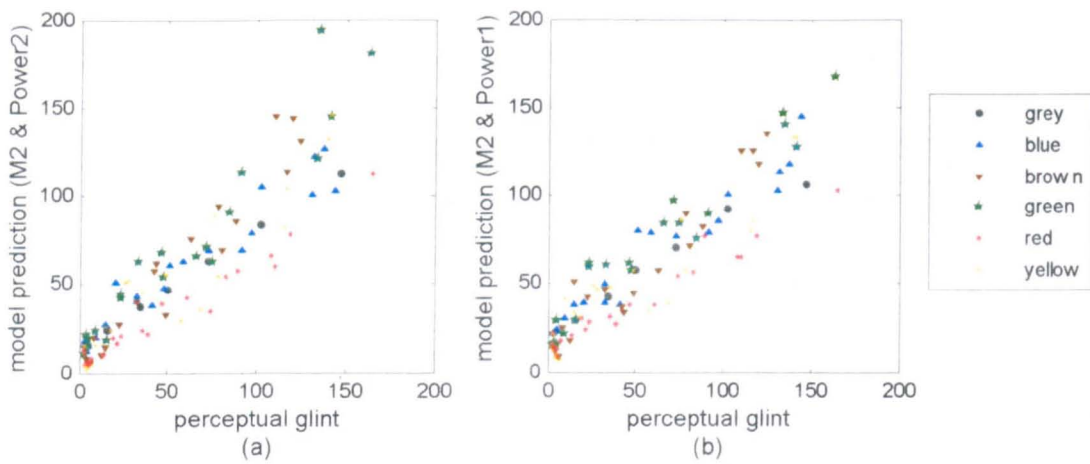


Figure 5-30: The model predictions against the perceptual glint. (a) The performance of the model incorporated with $M2$ and $Power2$ based on the luminance value obtained from the linear characterisation model. (b) The performance of the model incorporated with $M2$ and $Power1$ based on the luminance value obtained from the polynomial characterisation model.

The predictions of the model with $M2$ and $Power1$ (applied to the samples associated with the polynomial characterisation model) and the scale values were plotted separately for each colour of the samples with the indication of the associated correlation value of R^2 in Figure 5-31. Black dashed lines in graphs are linear regression lines between the model predictions and the scale values of all samples. Red lines in each graph are linear regression lines of each plotted data set. It can be seen the high correlations between the model predictions and the perceptual glint for each coloured samples. The R^2 values of 0.97, 0.93, 0.93, 0.94, 0.96 and 0.85 were obtained for the grey, blue, brown, green, red and yellow samples. The differences between the black dashed lines and the red lines indicate the dispersion of the tendency of each coloured sample from that of all the samples. These dispersions were different depending on the colours of the samples, however, the degree of the differences was quite small. Therefore, although it is not perfect, it can be concluded that there are no much differences in ability of the model to predict the perceptual glint depending on the colour of the samples. The same tendency was found from the observer variability (see Section 5.2.4.2). The observer variability did not seem to depend on the colours of the samples. It verifies that the luminance channel alone is sufficient for prediction glint not only for the grey samples but also for the coloured samples.

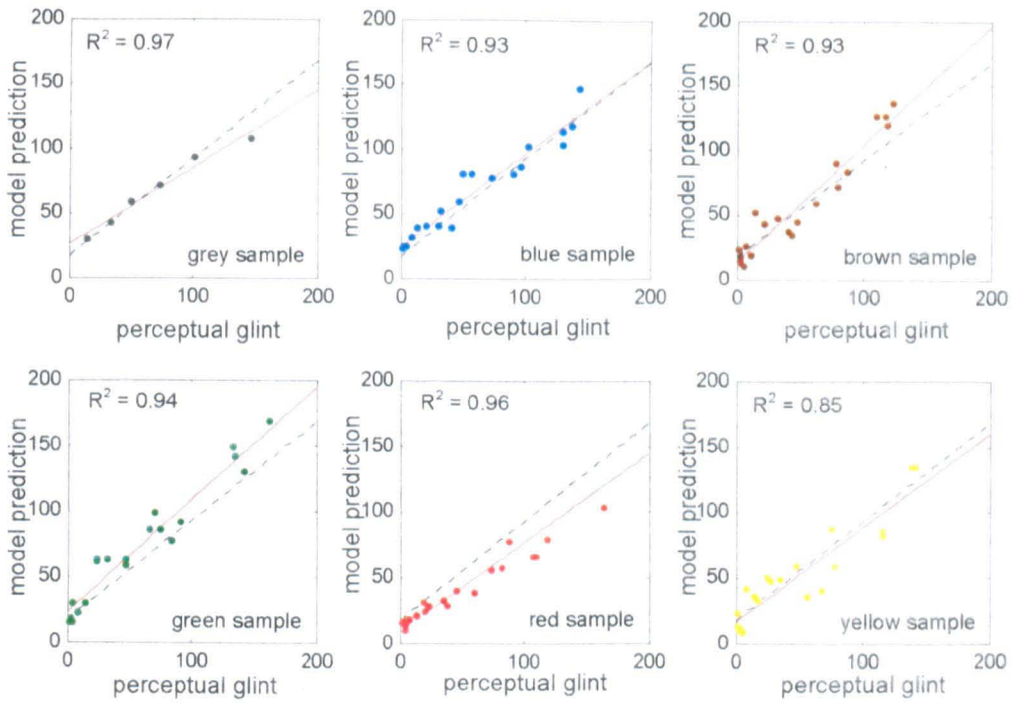


Figure 5-31: Model Predictions of the model incorporated with *M2* and *Power1* based on the luminance value obtained from the polynomial characterisation model plotted against the scale values for the samples in individual colour groups. Black dashed lines are linear regression lines to data (the scale values against the model predictions) for all samples. Red lines are liner regression lines of the plotted data.

The best performance of the model with the distribution-estimate method was the R^2 value of 0.85 and the RMS value of 0.116. They indicate slightly less accuracy in comparison with the best performance of the model with the iterative method which was the R^2 values of 0.94 and the RMSE values of 0.078. However, an advantage of the distribution-estimate method is a separate threshold to be derived for every image based upon the characteristics of that image. Hence, the model with the distribution-estimate method is considered to be sample-set independent except the final linearisation stage. Consequently, the distribution-estimate method is considered to be more robust. The final model consisting of *M2* and *Power1* with the distribution-estimate method is summarised in Equation 5-15.

$$Glint Model = \alpha \times \left(\frac{\sum (I(i, j) - p) \geq t_a}{m \times n} \right)^{\gamma} \quad \text{Equation 5-15}$$

$$t_a = 2p - l$$

where $I(i, j)$ is a pixel value (luminance value) of an image I at the pixel position of (i, j) ; $(i=1, 2, \dots, m)$ and $(j=1, 2, \dots, n)$, n and m represent the numbers of row and column in a

image, p and l are the mode and minimum value in an image, α and γ are model parameters. These parameter were $\alpha = 0.02$ and $\gamma = 0.68$ in this experiment.

5.5 Summary

In this chapter, the glint of the metallic-coating panels was characterised. The appearance of the glint is strongly affected by illumination and viewing geometry as well as other visual texture. Therefore, it was essential to give a strict definition of the observation geometry in order to specify the phenomenon of the glint. As a result of the visual assessment by observers, a suitable geometry was found to be using directional illumination such that an aspecular angle of about 71° . This was disagreed with the results of the glint assessment carried out at Akzo Nobel (see Section 5.2.5). They found that the glint was better observed with the observation angle close to a specular angle. However, at the same time, the study in Akzo Nobel revealed that the glint could be scaled with equivalent accuracy at different angles.

The glint of the metallic-coating samples was visually scaled by observers. As a result, the obtained observer variability suggests the high reliability of the scaled values of the perceptual glint for the samples and of the adequacy of the illumination and viewing condition for glint assessment. Also, it was found that that there were no significant differences in the variability between the samples having different colours.

In this chapter, the computational model was also developed to predict the perceptual glint from a digital image. An HDR imaging method was used to obtain useful images of metallic-coating samples containing aluminium flakes, since a LDR image was not sufficient to capture the glint effect. The HDR images were created based on the LDR images captured at two exposures and the spectral properties of a series of greyscale patches. The HDR images created had a capacity to cover the full range of the scene intensity of the samples whose maximum intensity was several times greater than the intensity of the white patch in ColorChecker DC seen under the experimental conditions used. This justified the use of the HDR approach in this study. The HDR values were then transformed to the CIE XYZ tristimulus values using a characterisation model. In a manner similar to that used in the coarseness model in Chapter 3, only a luminance channel (the Y tristimulus value) of the HDR image was incorporated for the glint feature extraction. The analysis of the luminance channel of the images revealed that the contrast information in the images were more relevant to

the glint than the absolute intensity. Then, the pixels in the image were segmented into two regions: spots (bright sparkles) and background (surrounds of the sparkles). Two approaches, an iterative method and a distribution-estimate method, were applied to find a threshold for segmenting. As a result of applying the various statistical analyses to the segmented image for a glint feature extraction, it was found that the perceptual glint was correlated to the pixels which belonged to the spots but the spatial distribution of these pixels could be ignored.

The glint values predicted using the model with the iterative method provided a better correlation with the perceptual glint than that using the model with the distribution-estimate method. However, there is a shortcoming for the iterative method. A single threshold value is used, but it is computationally intensive to derive, and it is not obvious that it would be applicable for samples that were not in the sample set from which it was derived. Therefore, this method is considered to be sample-set dependent. On the other hand, the distribution-estimate method allowed a separate threshold to be derived for every image based upon the characteristics of that image. Derivation of the threshold is computationally easy but relies upon certain assumptions that may not, in practice, be valid. The performance of the distribution-estimate method is slightly worse than the iterative method and most likely this is because the underlying assumptions are not met, however, the method is considered to be more robust whereas there is serious concern that the iterative method could not be relied on for samples that were not used in this study.

Chapter 6

Conclusions

6.1 Overview

In the automobile industry, there is considerable interest to be able to make quantitative measurements of appearance and to use digital images of the coating products, as perceived by customers, in order to improve efficiency in product development, product/quality control and communication. The overall appearance can be thought of as a combination of several different attributes such as colour, gloss, texture, *etc.* The present study focused only on modelling the visual texture, coarseness and glint of the metallic-coating panels, and to reproduce a digital image preserving the coarseness appearance of the panels. The investigation was carried using a subjective approach (perceptual scaling of coarseness and glint), and an objective approach (the derivation of computational models for the predicting the appearance attributes and the image reproduction of appearance). The major findings are summarised in the following sections.

6.1.1 Assessing Coarseness using Physical Samples

An experimental method for assessing the coarseness of metallic-coating panels was proposed. It was found that the coarseness could best be assessed under diffuse illumination. The observer accuracy obtained from the visual assessment of perceptual coarseness using scaling methods was less accurate in comparison with the coarseness assessment carried out in the similar conditions at Akzo Nobel (Kirchner and Kieboom *et al.*, 2007). This difference perhaps originated from the different observers involved. While naive observers participated in the present study, professional, and hence considerably more experienced observers, carried out the assessment at Akzo Nobel. Another difference between these assessments was that the present study used the 1-9 categorical judgement scaling method with only one reference sample, but Akzo Nobel employed a 0-9 categorical judgement scaling method with eight reference samples. Thus, it may be possible to improve the observer accuracy using more than one reference sample. The high value of observer accuracy obtained in both experiments suggests the adequacy of the diffuse illumination condition for coarseness assessment and that the perceptual coarseness can be reliably assessed by observers.

6.1.2 Computational Model for Coarseness Prediction

The present study proposed a computational model capable of predicting the perceptual coarseness of the metallic-coating panels, based on the information from a digital image of those panels. Figure 6-1 presents a flowchart of the main framework for the coarseness model. The model required an image of the metallic-coating panel, the physical size of the panel, the viewing distance and the white point for the observation condition as input. After transformation from the RGB values of each pixel in the image to the corresponding CIE XYZ tristimulus values, and then to the LMS values, the image was separated into a luminance and two chromatic channels. Only the luminance channel was utilised for next processes. In order to measure the contrast, the mean luminance value of all pixels in the image was subtracted from each pixel value. The Fourier energy was computed as a measure of the contrast and then the energy was weighted by the contrast-sensitivity function (CSF) to take into account the characteristics of the human visual system. The sum of these processed pixel values was then normalised using the mean luminance value for incorporating with another characteristic of the CSF, *i.e.*, Weber's law. Finally, the coarseness prediction was obtained by taking the logarithm of such a sum after normalisation by the number of the pixels in the image.

The advantages of this model are that it is computationally inexpensive and the input parameters are related to the viewing conditions. Unlike the conventional texture models, such as SGLDM, GLDM, NGLDM and grey level run length (see Section 2.6), no parameters needed to be estimated. The performance of the model was investigated by comparing model predictions with the perceptual coarseness scaled by observers. The model gave excellent performance in terms of the accuracy in predicting the visual results. There were variations in the model performance depending on the colour of the samples, for example, the performance for the blue and purple colour samples was generally poorer than that of the green samples. This might be caused by the fact that the model only considered the information of the luminance channel and ignored that of the chromatic channels. However, because of the limited number of samples, *i.e.*, the available colours and coarseness levels of the metallic-coating panels supplied by Azko Nobel, it was not possible to determine the importance of the chromatic channels for coarseness analysis.

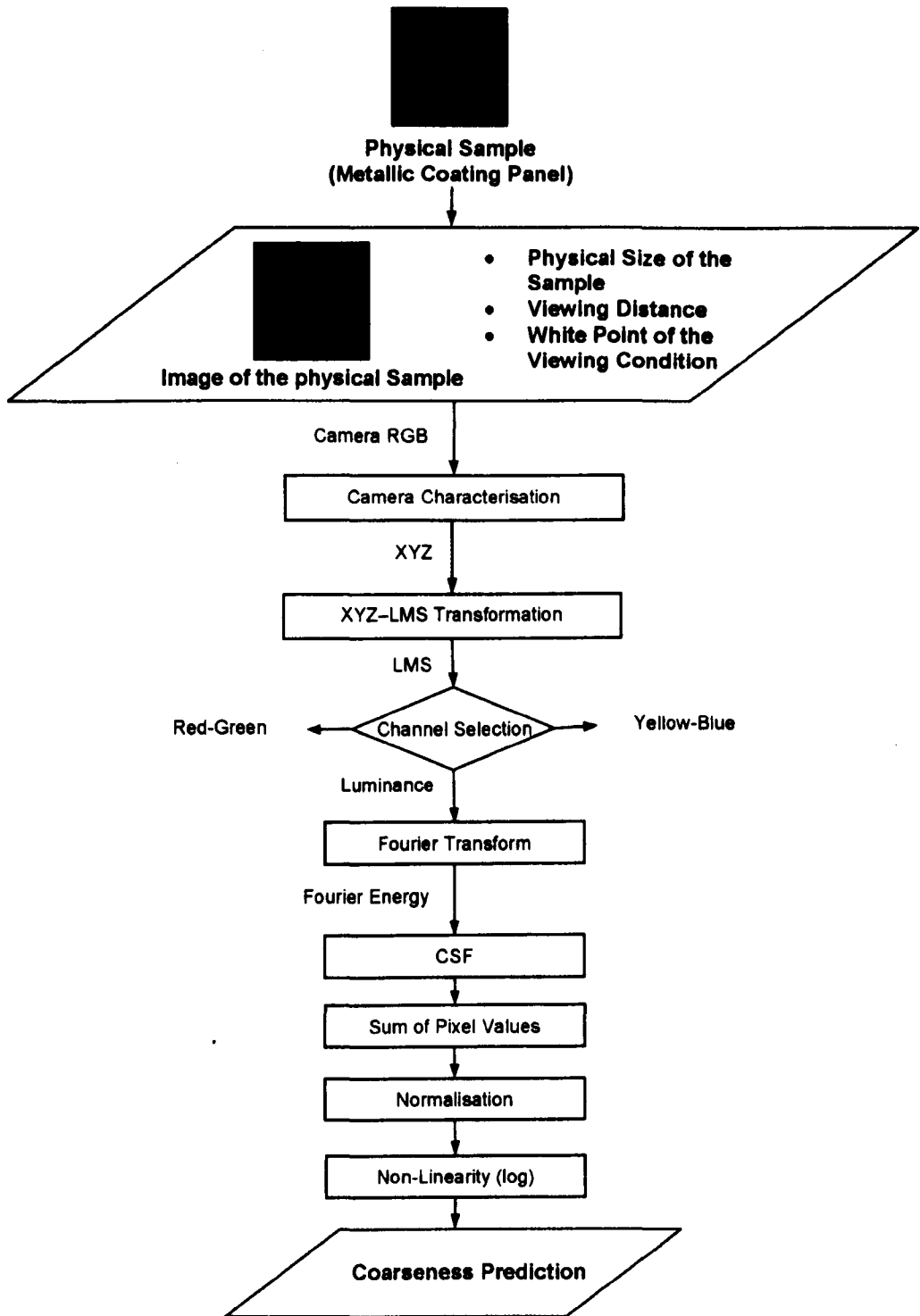


Figure 6-1: A flowchart of the main framework for the developed coarseness model.

6.1.3 Assessing Coarseness using a Display

The coarseness appearance of the metallic-coating panels was reproduced on an LCD and their perceptual coarseness was scaled by observers using the similar scaling method as used for the metallic-coating panels (physical samples). The sample images were generated based on the images used as input for the coarseness model that was used to predict the coarseness perceived on the physical samples. The images were required to be captured with high resolution to capture a suitable level of detail. Consequently, it was not possible to display the image with an actual size that was equivalent to the size of the physical sample (8×8 cm) because of the resolution limitation of the LCD used in this experiment. Therefore, two optimal conditions were chosen. One kept the experimental conditions used in the assessments for the physical samples in the same ratio. This resulted in the physical size of the images of the samples being much larger, and the viewing distance thus greater, than that used for the experiment with the physical samples, but the angular display size was kept at a similar value. The perceptual coarseness in this condition was smaller than that observed on the real samples. The second condition used the same angular display size and had the similar physical image size and viewing distance to the conditions used for scaling the physical samples. Thus, this resulted in the presentation of only a part of the image of the physical sample. However, the perceptual coarseness in this condition appeared closer to that observed on the real samples.

As results of the perceptual coarseness scaling of the image samples, observer variability obtained at these two conditions was similar and it was also similar to that obtained from the coarseness assessment using the physical samples. This indicates that the observers assessed the coarseness with equal precision. The comparison of the perceptual coarseness scaled using these three conditions revealed that although the conditions were different, there were no particular differences in the relative scaled perceptual coarseness. This suggests that for assessing the equally distributed fine detail over a sample like the metallic-coating panel used in the experiment, the area presented to the observers is not so important. Observers tend to focus on only a part of the sample and not on the whole sample. Although the quality of the reproduced images was proved in terms of the perceptual coarseness, a significant colour shift was found. The lightness of the images was relatively preserved compared with the errors that occurred in the values of chroma and hue. These results suggest the importance of lightness for the perception of coarseness rather than chroma and hue, and also support

the coarseness model that uses only the luminance channel to predict perceptual coarseness. In fact, the model predictions for each condition all agreed well with the scaled values of perceptual coarseness.

6.1.4 Assessing Glint using Physical Samples

The present study was also aimed to characterise the perceptual glint of the metallic-coating panels. As well as the coarseness, the perceptual glint of the coatings varies with the illumination and viewing geometry. Therefore, the illumination and viewing geometry for which the glint was most obvious was determined by visual assessment. A suitable geometry was found to be using directional illumination such that an aspecular angle of about 71° . The perceptual glint of the metallic-coating panels was scaled under the selected conditions using the magnitude estimation method with a single reference sample. As a result, the high value of observer accuracy and repeatability were obtained. These suggest the reliability of the scaled values of the perceptual glint for the panels and of the adequacy of the illumination and viewing condition for glint assessment.

6.1.5 Computational Model for Perceptual Glint Prediction

A computational model was developed to predict glint using images of the metallic-coating panels. Figure 6-2 presents a flowchart of the main framework for the developed glint model. The model employed an HDR imaging system in order to capture the full range of the glint information of the metallic-coating panels observed under the directional illumination. The HDR images were created based on the LDR images captured at two exposures and the spectral properties of a series of greyscale patches. Each HDR pixel value was basically obtained by taking a pixel value from either one of the LDR images representing the two exposures or the average of the two LDR images at the two exposures. This simple algorithm allowed for the creation of the HDR images with suitable high precision. Then the CIE XYZ tristimulus values were transformed from the HDR values using a camera characterisation model. In a manner similar to that used in the coarseness model, only a luminance channel (the Y tristimulus value) of the HDR image was incorporated for the glint feature extraction. In order to analyse the contrast in the image rather than the absolute values, the mode of the luminance channel of the image was subtracted from each pixel value. Then a

thresholding technique was applied to segment pixels in the image into two regions: spots and background. In this study, two segmentation approaches, an iterative method and a distribution-estimate method, were applied. Various statistical approaches to the segmented image revealed that the sum of the pixel values that exceeded a certain threshold, *i.e.*, the pixel values that belonged to the spots, was correlated with the perceptual glint. For the model incorporated with the distribution-estimate method, such a sum, linearised using a power function, was found to be proportional to the perceptual glint. For the model incorporated with the iterative method, such a sum normalised by the mode of individual images for incorporating with Weber's law, and then linearised using a power function, was proportional to the perceptual glint. Although the model with the iterative method provided better predictions than the model with the distribution-estimate method, there is a shortcoming in the iterative method. The iterative method may not be applied to samples if colour distribution and glint levels are significantly different from samples which are used to derive an appropriate threshold value. Hence, this method is considered to be sample-set dependent. On the other hand, the distribution-estimate method allows a separate threshold to be derived for each image based on the characteristics of that image and the derivation of the threshold is computationally easy. Hence, this model can be considered to be sample-set independent. Consequently, the distribution-estimate method is considered to be more robust.

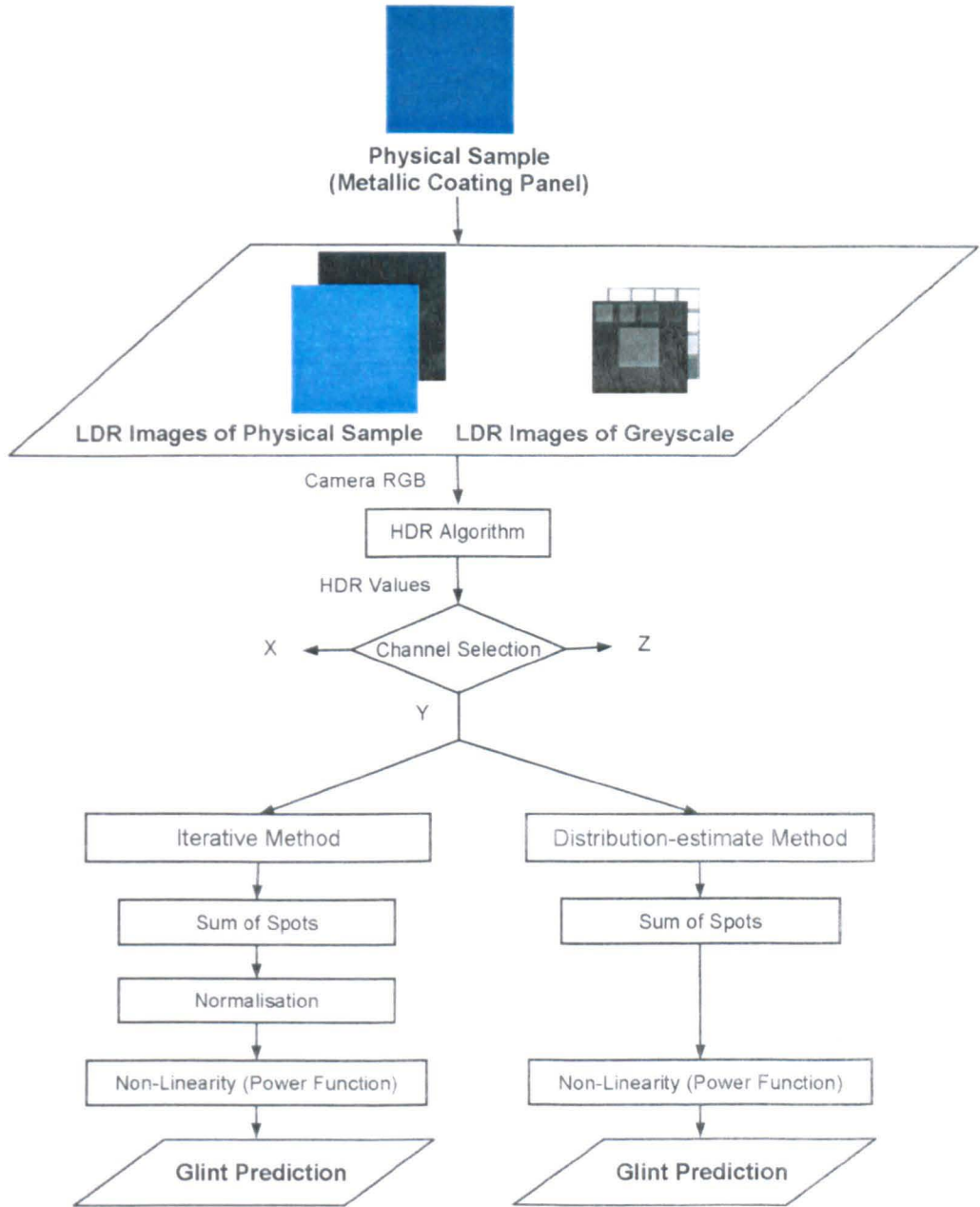


Figure 6-2: A flowchart of the main framework for the developed glint model.

6.2 Discussion and Future Work

In order to develop a system for the measurement of the appearance of a metallic coating, two important appearance attributes, coarseness and glint, were studied. One of the major limitations of the present study was in the sample sets used for the investigation. The experiments were carried out using samples with only limited colours and coarseness or glint levels; variations in existing metallic coatings are significantly larger and new products are frequently introduced. Because of these limitations in the samples used, it was not possible to investigate the details of the influence of the colour on perceptual coarseness and glint. Although the developed models for both attributes provided satisfactory performance by only taking into account the luminance channel of the image, to include the chromatic channels might further improve the model performance.

The developed models were both based on the image captured by a digital camera and the associated computation was inexpensive. Therefore, the concept has practical application to instruments for appearance measurement. For example, a BYK-mac® has recently been developed by BYK-Gardner GmbH in close collaboration with Akzo Nobel and this instrument measures colour using various geometries as well as two appearance attributes, coarseness and sparkle.

The relative visualisation of the coarseness appearance was found to be possible. It was however, not possible to achieve absolute visualisation due to technical limitations. Although the digital camera allowed the capture of images at high resolution, the display resolution was not high enough. Also, in this study, only the visualisation of the coarseness was attempted; not the glint. This was because the high dynamic range of the glint information was beyond the dynamic range of the LCD used. It would be possible to overcome these technical limitations using a higher resolution HDR display. Recent high-end technologies make it possible to achieve, for example, a 22 inch LCD with a resolution of 3840×2400 (produced by Mondale IT Solutions Ltd), and an HDR display providing a contrast ratio of 200000:1 with a luminance of over 3000 cd/m^2 (from Dolby). It is predicted that high resolution HDR displays will become popular in the near future and this will allow better simulation of visual appearance. Other than these limitations, the accuracy of the colour in reproduced image needs to be improved. The reproduced images were evaluated perceptually in terms of the coarseness, but not in term of the perceptual colour. Since the appearance

is the interaction of many attributes, for accurate visualisation of total appearance, it is necessary to reproduce the colour dimension accurately as well as the spatial attributes.

In the present study, the reproduction error of the images was measured colorimetrically using CIELAB colour difference pixel by pixel. As S-CIELAB suggests (Section 2.3.7), the colour difference values given from the computation of CIELAB colour difference for a non-uniform colour image tend to overestimate the perceived colour difference. Therefore, it can be thought that a better measure of the perceived colour difference is to use the S-CIELAB method rather than the CIELAB equation. However, in the present study, the magnitude of the perceptual colour difference was not the focal point. The relative comparison of the systematic error that occurred in the reproduction process was focused and evaluated. Therefore, it was convenient to use a consistent measure, the CIELAB colour difference, for all processes such as the camera characterisation, the monitor characterisation and the reproduced images. For the visualisation of total appearance, it would be necessary to evaluate the perceptual colour difference using a more accurate method such as S-CIELAB, and it would be ideal to reproduce the images of the metallic coatings accurately in terms of not only perceptual coarseness but also colour, glint and other attributes.

In the present study, the attributes of coarseness and glint were investigated individually. Since different sets of the samples were used for the coarseness and glint investigations, it was not possible to compare and find the correlation between these two attributes. However, it would be interesting to know the relationship between two attributes. It is however, possible that these two attributes may be considered insufficient to describe the total appearance of the coatings. Colour has also been studied as an important attribute (Section 2.5.1.1), but how many more attributes need to be characterised to measure the total appearance of the metallic coatings is still unknown. Therefore, another area of future work should be focused on other appearance attributes which were not investigated in this study.

References

- Akzo-Nobel (2004). Akzo-Nobel Internal Report.
- Akzo-Nobel (2004-2006). Private Communication.
- Alman DH (1984). Three Direction Measurements for Characterization of a Surface Containing Metallic Particles. 4479718 USP, E. I. Du Pont de Nemours and Company, USA.
- Arai H & Baba G (2005). Micro-brilliance of Anisotropic Paint Surfaces. In *Proceedings of AIC Colour 2005, the 10th Congress of the International Colour Association*, Granada, Spain, 813-816.
- ASTM (2001). ASTM E284-07: Standard Terminology of Appearance: Annual Book of ASTM Standards: Textiles. **07-01**.
- ASTM (2003). ASTM E2194-3: Standard Practice for Multiangle Color Measurement of Metal Flake Pigmented Materials (in development by Subcommittee: E12.12).
- ASTM (2006). ASTM D1729-96: Standard Practice for Visual Appraisal of Colors and Color Differences of Diffusely-Illuminated Opaque Materials. **06-01**, 217-220.
- Atchison DA & Smith G (2000). *Optics of the Human Eye*. Butterworth-Heinemann Ltd., Massachusetts, USA, 5,9,12.
- Baba M, Miura S, Mukunoki M & Asada N (2005). Reflectance Estimation of Sparkle in Metallic Paints. In *Proceedings of International Conference on Computer Graphics and Interactive Techniques*, California, USA.
- Barten PGJ (1990). Evaluation of Subjective Image Quality with the Square-Root Integral Method. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, **7** (10), 2024-2031.
- Bartleson JC (1979). Changes in Color Appearance with Variations in Chromatic Adaptation. *Color Research and Application*, **4** (3), 119-138.
- Bartleson JC (1984). Vol. 5 Visual Measurements. In *Optical Radiation Measurements*. ed. Bartleson J, C. & Grum F, Academic Press, New York, USA, 473-493, 491-505.
- Bayer BE (1976). Color Imaging Array. 3971065 USP, Eastman Kodak Company, USA.
- Beer A (1852). Bestimmung der Absorption des rothen Lichts in farbigen. *Ann Physik Chem*, **86** (2), 78.
- Benschawel T (1992). Color Vision, Perception and Measurement – Colorimetry of Self-Luminous Displays. In *Color in Electronic Displays* ed. Widdel H & Post DL, Springer, New York, USA, 85-86.
- Berns RS (1996). Methods for Characterizing CRT Displays. *Displays*, **16** (4), 173-182.

- Berns RS, Grozynski ME & Motta RJ (1993). CRT Colorimetry. Part II: Metrology. *Color Research and Application*, **18** (5), 315-325.
- Berns RS, Motta RJ & Grozynski ME (1993). CRT Colorimetry. Part I: Theory and Practice. *Color Research and Application*, **18** (5), 299-314.
- Besold R (1990). Metallic Effect - Characterization, Parameter and Methods for Instrumentally Determination. *Die Farbe*, **37**, 79-85.
- Boettner EA & Wolter JR (1962). Transmission of the Ocular Media. *Investigative Ophthalmology & Visual Science*, **1**, 776-783.
- Bouzit S & MacDonald L (2001). Sharpness Enhancement through Spatial Frequency Decomposition. In *Proceedings of the Image Processing, Image Quality, Image Capture Systems Conference (PICS-01)*, Quebec, Canada, 377-381.
- Cambridge (2005). *Cambridge Advanced Learner's Dictionary*. 2nd Edition. Cambridge University Press, UK.
- Campbell FW & Robson JG (1968). Applications of Fourier Analysis to the Visibility of Gratings. *Journal of Physiology*, **181**, 576-593.
- Carlson C (1982). Sine-wave Threshold Contrast-sensitivity Function: Dependence on Display Size. *RCA Review*, **43**, 675-683.
- Castleman KR (1996). *Digital Image Processing*. Prentice Hall Inc., NJ, USA.
- CAVE (2007). High Dynamic Range Imaging: Multiple Exposures. http://www1.cs.columbia.edu/CAVE/projects/hdr_me/hdr_me.php. Columbia University, Computer Vision Laboratory.
- Chen CH, Pau LF & Wang PSP (1999). *Handbook of Pattern Recognition & Computer Vision*, 2nd Edition, World Scientific Publishing Co. Pte. Ltd., Singapore.
- Cheung TLV (2004). Approach to Colour Camera Characterization. *PhD Thesis*, University of Leeds, UK.
- Chou W (2003). Evaluation of Lightness Differences and Metallic Colour Differences. *PhD Thesis*, University of Derby, UK.
- CIE (2004a). CIE Technical Report, A Colour Appearance Model for Colour Management Systems: CIECAM02. *CIE Publication 159*.
- CIE (2004b). CIE Technical Report, Colorimetry. *CIE Publication 15*, 3rd Edition.
- Connors RW & Harlow CA (1980). A Theoretical Comparison of Texture Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **2** (3), 204-222.

- Cui G, Luo MR, Rigg B & Li W (2001). Colour-Difference Evaluation using CRT Colours. Part I: Data Gathering and Testing Colour Difference Formulae. *Color Research and Application*, **26** (5), 394-402.
- Curcio CA, Allen K, Sloan K, Lerea C, Hurley J, Klock I & Milam A (1991). Distribution and Morphology of Human Cone Photoreceptors Stained with Anti-blue Opsin. *Journal of Comparative Neurology*, **312**, 610-624.
- CVRL (2007). Color Matching Functions. <http://www.cvrl.org/cmfs.htm>. Colour Vision Research Laboratories, Institute of Ophthalmology, University College London.
- Day EA, Taplin L & Berns RS (2004). Colorimetric Characterisation of a Computer-Controlled Liquid Crystal Display. *Color Research and Application*, **29** (5), 365-373.
- Dean A & Voss D (1999). *Design and Analysis of Experiments*. Springer Verlag, New York, USA.
- Debevec PE & Malik J (1997). Recovering High Dynamic Range Radiance Maps From Photographs. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, 369-378
- Đurikovič R (2003). Simulation of Sparkling and Depth Effect in Paints. In *Proceedings of the 19th ACM Spring Conference on Computer Graphics – SCCG2003*, Budmerice, Slovakia, 193-198.
- Engeldrum PG (2000). *Psychometric Scaling: A Toolkit for Imaging Systems Development*. Imcotek Press, Massachusetts, USA, 43-52, 93-157.
- Ershov S, Kolchin K & Myszkowski K (2001). Rendering Pearlescent Appearance Based on Paint-composition Modelling. *Computer Graphics Forum*, **20** (3), C221-C238.
- Fattal R, Lischinski D & Werman M (2002). Gradient Domain High Dynamic Range Compression. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, USA, 249-256
- Galloway MM (1975). Texture Classification using Grey Level Run Length. *Computer Graphics and Image Processing*, 172-179.
- Gescheider GA (1997). *Psychophysics: The Fundamentals*. 3rd Edition. Lawrence Erlbaum Associates Inc., New Jersey, USA, 215-216.
- Gibson JE & Fairchild MD (2000). Colorimetric Characterization of Three Computer Displays (LCD and CRT). *Munsell Color Science Laboratory Technical Report*.
- Gibson JJ (1950). *The Perception of the Visual World*. Houghton-Mifflin, MA, USA.
- Gonzalez RC & Woods RE (1992). *Digital Image Processing*. Addison Wesley Publishing Company, USA, 81-92.

- Gonzalez RC, Woods RE & Eddins SL (2004). *Digital Image Processing Using MATLAB*. Prentice Hall, New Jersey, USA, 359-362.
- Green P (2002a). Characterizing Hard Copy Printers. In *Colour Engineering*. ed. Green P & MacDonald L, John Wiley & Sons Ltd., West Sussex, UK, 221-246.
- Green P (2002b). Overview of Characterization Models. In *Colour Engineering*. ed. Green P & MacDonald L, John Wiley & Sons Ltd., West Sussex, UK, 127-140.
- Gremer TA & Nadel ME (2001). Modeling the Appearance of Special Effect Pigment Coatings. In *Proceedings of SPIE: Surface Scattering and Diffraction for Advanced Metrology*, CA, USA, 77-86.
- Guild J (1931). The Colorimetric Properties of the Spectrum. *Philosophical Transactions of the Royal Society of London*, **A230**, 149-187.
- Han B (2006). Texture Simulation of Coloured Objects on CRT Displays. *PhD Thesis*, University of Derby, UK.
- Han B, Luo MR & Kirchner EJJ (2005a). Assessing Colour Differences for Automobile Coatings using CRT Colours Part I: Evaluating Colour Difference of Solid Colours. In *Proceedings of AIC Colour 2005, the 10th Congress of the International Colour Association*, Granada, Spain, 579-582.
- Han B, Luo MR & Kirchner EJJ (2005b). Assessing Colour Differences for Automobile Coatings using CRT Colours Part II: Evaluating Colour Difference of Textured Colours. In *Proceedings of AIC Colour 2005, the 10th Congress of the International Colour Association*, Granada, Spain, 583-586.
- Haralick RM (1979). Statistical and Structural Approaches to Texture. In *Proceedings of the IEEE*, **67** (5) 786-804.
- Haralick RM, Shanmugam K & Dinstein I (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-3** (6), 610-621.
- Hardeberg JY (1999). Acquisition and Reproduction of Colour Images: Colorimetric and Multispectral Approaches. *PhD Thesis*, École Nationale Supérieure des Télécommunications, France.
- Hendee WR & Wells PN (1997). *The Perception of Visual Information*. 2nd Edition. Springer, New York, USA, 57-83.
- Hong G, Luo MR & Rhodes PA (2001). A Study of Digital Camera Colorimetric Characterization bases on Polynomial Modeling. *Color Research and Application*, **26** (1), 78-84.
- HORIBA JobinYvon (2007). Application Note.
<http://www.jobinyvon.com/usadivisions/TFilms/applications/se-03.pdf>

- Hung P-C (1993). Colorimetric Calibration in Electronic Imaging Devices using a Look-up Tables Model and Interpolations. *Journal of Electronic Imaging*, **2** (1), 53-61.
- Hunt RWG (1998). *Measuring Colour*. 3rd Edition. Fountain Press Kingston-upon-Thames, UK, 19, 99-101.
- Hunter R (1977). Applications of Appearance Measurements in the Coatings Industry. *Journal of Coatings Technology*, **49**, 87-91.
- Hunter RS (1987). *The Measurement of Appearance*. John Wiley & Sons Inc., New York, USA, 63-67, 89.
- IEC (1998). IEC/3WD 61966-2-1: Colour Measurement and Management in Multimedia Systems and Equipment – Part 2-1: Default RGB Color Space – sRGB. *International Electrotechnical Commission*.
- Imura K (2006). Optical Measuring Apparatus, Illumination System, and Light Detecting System. 20060109474 USP, USA.
- Ippolito JA (2002). *Understanding Digital Photography*. Delmar Learning, NY, USA, 31-64.
- Iwata M, Okajima K & Ujike H (2001). Change in Visual Contrast Sensitivity with Age Dependends on Illuminance Level. *Journal of the Illuminating Engineering Institute of Japan*, **85** (5), 352-360.
- Johnson GM & Fairchild MD (2003). Visual Psychophysics and Color Appearance. In *Digital Color Imaging Handbook*. ed. Sharma G, CRC Press Inc., FL, USA, 124, 125.
- Johnson GM & Montag ED (2005). Size Matters: The Influence of Viewing Distance on Perceived Spatial Frequency and Contrast. In *Proceedings of the 13th Color Imaging Conference: Color Science and Engineering: Systems, Technologies, Applications*, Arizona, USA, 339-343.
- Johnson T (1996). Methods for Characterizing Colour Scanners and Digital Cameras. *Displays*, **16** (4), 183-191.
- Kang H, Butler C, Yang Q, Sacerdotti F & Benati F (2000). Appearance Measurement System using Fuzzy Logic. In *Proceedings of the 17th IEEE Instrumentation and Measurement Technology Conference*, 1124-1129.
- Kang HR (1997). *Color Technology for Electronic Imaging Devices*. SPIE Optical Engineering Press, 55-100.
- Katoh N, Deguchi T & Berns R, S. (2001a). An Accurate Characterization of CRT Monitor (I) Verifications of Past Studies and Clarifications of Gamma. *Optical Review*, **8** (5), 305-314.
- Katoh N, Deguchi T & Berns RS (2001b). An Accurate Characterization of CRT Monitor. (II). Proposal for an Extension to CIE Method and Its Verification. *Optical Review*, **8** (5), 397-408.

- Kirchner E, van den Kieboom G-J, Njo L, Super R & Gottenbos R (2007). Observation of Visual Texture of Metallic and Pearlescent Materials. *Color Research and Application*, **32** (4).
- Kitaguchi S, Westland S, Owens H, Luo MR & Pointer MR (2004). Surface Texture - A Review. *National Physical Laboratory Report DQL-OR 006*, Teddington, MdIx, UK.
- Kothari RC (2005). *Research Methodology: Methods and Techniques*. 2nd Edition. New Age Publishers, India, 291-298.
- Krawczyk G, Goesele M & Seidel H-P (2005). Photometric Calibration of High Dynamic Range Cameras. *MPI Informatik Technical Report MPI-I-2005-4-005*, Germany.
- Kruizinga P, Petkov N & Grigorescu SE (1999). Comparison of Texture Features based on Gabor Filters. In *Proceeding of the 10th International Conference on Image Analysis and Processing*, Venice, Italy, 142-147.
- Kuang J, Yamaguchi H, Johnson GM & Fairchild MD (2004). Testing HDR Image Rendering Algorithms. In *Proceedings of the 12th Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications* Arizona, USA, 315-320.
- Kubelka P & Munk F (1931). Ein Beitrag zur Optik der Farbanstriche. *Z. Tech. Physik*, **12**, 593-601.
- Kwak Y & MacDonald L (2001). Accurate Prediction of Colours on Liquid Crystal Displays In *Proceedings of the 9th Color Imaging Conference: Color Science and Engineering: Systems, Technologies, Applications*, Arizona, USA, 355-359.
- Lambert JH (1760). *Photometria sine de mensura et gradibus luminis colorum et umbrae*. *Eberhardi Klett*.
- Larson GW, Rushmeier H & Piatko C (1997). A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes. *IEEE Transactions On Visualization And Computer Graphics*, **3** (4), 291-306.
- Laws KI (1980). Textured Image Segmentation. *PhD Thesis*, University of Southern California, USA.
- Lee H-C (2005). *Introduction to Color Imaging Science*. Cambridge University Press, Cambridge , UK, 49, 294-312, 305, 306, 333, 494-511, 532-536.
- Lew MS (2001). *Principles of Visual Information Retrieval*. Springer, UK, 64.
- LG Philips (2007). <http://www.lgphilips-lcd.com/>.
- Li C (2008). Private Communication.
- Litwiller D (2005). *CCD vs. CMOS: Maturing Technologies, Maturing Markets*. Laurin Publishing, 54-58.
- Lloyd CD (2007). *Local Models for Spatial Analysis*. CRC Press Inc., FL, USA, 33-34.

- Luce DR & Krumbhansl CL (1988). Measurement, scaling, and psychophysics. In *Stevens' Handbook of Experimental Psychology: Perception and Motivation*, 2nd Edition, vol. 1. ed. Atkinson RC, Herrnstein RJ, Lindzey G & Luce DR, John Wiley & Sons Inc., NY, USA, 49, 54-59.
- Luce DR & Suppes P (2002). Representational Measurement Theory. In *Stevens' Handbook of Experimental Psychology: Methodology in Experimental Psychology*, 3rd Edition, vol. 4. ed. Pashler HE & Wixted J, John Wiley & Sons Inc., New York, USA, 14, 24-26.
- Luo MR (2002). *Text Book: Colour Science for MSc Imaging Science*. University of Derby, UK.
- Luo MR, Clarke AA, Scrivener SAR & Tait CJ (1991). Quantifying Colour Appearance. Part I. LUTCHI Colour Appearance Data. *Color Research and Application*, **16** (3), 166-180.
- Luo MR, Cui G & Li C (2001). *British Patent: Apparatus and method for measuring colour (DigiEye system)*. Derby University Enterprise Limited, Derby, UK.
- MacLeod DIA & Boynton RM (1979). Chromaticity Diagram Showing Cone Excitation by Stimuli of Equal Luminance. *Journal of the Optical Society of America*, **69** (8), 1183-1186.
- Malacara D (2002). *Color Vision and Colorimetry: Theory and Applications*. SPIE: The International Society for Optical Engineering, WA, USA, 132.
- Matlin MW & Foley HJ (2007). The Visual System. <http://www.skidmore.edu/~hfoley/Perc3.htm>.
- Maxwell JC (1980). On the Theory of Compound Colors and Relations of the Colors of the Spectrum. *Philosophical Transactions*, **150**, 57-84 [Reprinted with commentary by Qasim Zaidi (1993) in: *Color Research and Application*, **18**, 1270-1287].
- McCamy CS (1996). Observation and Measurement of the Appearance of Metallic Materials. Part I. Macro Appearance. *Color Research and Application*, **21** (4), 292-304.
- McCamy CS (1998). Observation and Measurement of the Appearance of Metallic Materials. Part II. Micro Appearance. *Color Research and Application*, **23** (6), 362-373.
- Merriam-Webster (2005). Dictionary and Thesaurus: Merriam-Webster Online. <http://www.m-w.com/dictionary/coarseness>.
- Meylan L (2006). Tone Mapping for High Dynamic Range Images. *PhD Thesis*. École Polytechnique Fédérale de Lausanne, France.
- Meylan L & Süsstrunk S (2006). High Dynamic Range Image Rendering With a Retinex-Based Adaptive Filter. *IEEE Transactions on Image Processing*, **15** (9), 2820-2830.
- Michelson AA (1927). *Studies in Optics*. University of Chicago Press, Chicago, USA.
- Miller N (2007). Lighting Design. <http://www.nmlightingdesign.com/topics/index.php>.

- Mitsunaga T & Nayar SK (1999). Radiometric Self Calibration. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, USA*, 374-380.
- Morovič J (1998). To Develop a Universal Gamut Mapping Algorithm. *PhD Thesis*, University of Derby, UK.
- Mullen KT (1985). The Contrast Sensitivity of Human Colour Vision to Red-green and Blue-yellow Chromatic Gratings. *Journal of Physiology*, **359**, 381-400.
- Nadal ME & Early EA (2004). Color Measurements for Pearlescent Coatings. *Color Research and Application*, **29** (1), 38-42.
- Nayar SK & Mitsunaga T (2000). High Dynamic Range Imaging: Spatially Varying Pixel Exposures. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, USA*, 472-479.
- Nicholls CA (2000). Visual Instrumental Characterisation of Special Effect Colours. *PhD Thesis*, University of Leeds, UK.
- O'Malley SM (2006). A Simple, Effective System for Automated Capture of High Dynamic Range Images. In *Proceedings of the Fourth IEEE International Conference on Computer Vision Systems (ICVS 2006)*, USA, 15.
- Ohanian PP & Dubes RC (1992). Performance Evaluation for Four Class of Texture Features. *Pattern Recognition*, **25** (8), 819-833.
- Ohta N (1982). Practical Transformations of CIE Color Matching Functions. *Color Research and Application*, **7** (1), 53-56.
- Osterberg G (1935). Topography of the Layer of Rods and Cones in the Human Retina. *Acta Ophthalmol*, **6**, 1-103.
- Owens HC (2002). Colour and Spatiochromatic Processing in the Human Visual System. *PhD Thesis*, University of Derby, UK.
- Packer O & Williams DR (2003). Light, the Retinal Image, and Photoreceptors. In *The Science of Color*, 2nd Edition. ed. Shevell SK, Elsevier, Oxford, UK, 41-97.
- Park J-Y, Jeon S-Y, O'mahony M & Kim K-O (2004). Induction of Scaling Errors. *Journal of Sensory Studies*, **19** (4), 261-271.
- Pointer MR (1980). The Concept of Colourfulness and its Use for Deriving Grids for Assessing Colour Appearance. *Color Research and Application*, **5** (2), 99-107.
- Pointer MR (2003). Measuring Visual Appearance - A Framework for the Future. *National Physical Laboratory Report COAM*, Teddington, Mdx, UK.

- Portilla J & Simoncelli EP (2000). A Parametric Texture Model Based on joint Statistics of Complex Wavelet Coefficients. *International Journal of Computer Vision*, **40** (1), 49-71.
- Prakash A, Karmes RA, Obetz JE, Schenk WB, Gallagher JP, Blase A & Rodrigues J (2005). Method of Characterization of Surface Coating Containing Metallic Flakes and Device Used Therein. 6952265 B2 USP, E. I. du Pont de Nemours and Company, USA.
- Purves D, Augustine GJ, Fitzpatrick D, Hall WC, Lamantia A-S, McNamara JO & White LE (2001). *Neuroscience*. 2nd Edition. Sinauer Associates Inc., MA, USA, <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=.0Y4lv4ycgt6m3Im-1-51KjOIAw>.
- Randen T & Husøy JH (1999). Filtering for Texture Classification: A Comparative Study. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **22** (4), 291-310.
- Reinhard E, Ward G, Pattanaik S & Debevec P (2006). *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. Morgan Kaufmann San Francisco, USA, 6, 118-121.
- Rhodes PA (2007). Private Communication.
- Richardson LF & Ross JS (1930). Loudness and Telephone Current. *Journal of General Psychology*, **3**, 288-306.
- Robertson MA, Borman S & Stevenson RL (2003). Estimation-theoretic Approach to Dynamic Range Enhancement Using Multiple Exposures. *Journal of Electronic Imaging*, **12** (2), 219-228.
- Rodrigues A (2004). Color Technology and Paint. In *Proceedings of AIC Color 2004 Color and Pints, Interim Meeting of the International Color Association*, Porto Alegre, Brazil, 103-108.
- Saris HJA, Gottenbos RJB & Houwelingen H, van (1990). Correlation Between Visual and Instrumental Colour Differences of Metallic Paint Films. *Color Research and Application*, **15** (4), 200-205.
- Saxby G (2002). *The Science of Imaging: An Introduction*. CRC Press Inc., FL, USA, 28.
- Sharma G (2002). LCDs Versus CRTs-Color-Calibration and Gamut Considerations. In *Proceedings of the IEEE*, **90** (4), 605-622.
- Shen H & Bie C (1992). Representation of Visual Texture Properties. In *Advances in Machine Vision: Strategies and Applications*. ed. Archibald C & Petriu E, World Scientific Publishing Co. Pte. Ltd., London, UK, 195-196.
- Siew LH, Hodgson RM & Wood EJ (1988). Texture Measures for Carpet Wear Assessment. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **10** (1), 92-105.
- Singh M & Singh S (2002). Spatial Texture Analysis: a Comparative Study. In *Proceedings of the 16th international conference on pattern recognition*, **1**, Quebec, Canada, 676-679.

- Speranskaya NI (1959). Determination of Spectrum Color Co-ordinates for Twenty-Seven Normal Observers. *Optics and Spectroscopy*, **7**, 424-428.
- Stevens SS (1953). On the Brightness of Lights and the Loudness of Sounds. *Science*, **118**, 576.
- Stevens SS (1961). To Honor Fechner and Repeal His Law. *Science*, **133**, 80-86.
- Stiles WS & Burch JM (1959). NPL Colour-Matching Investigation: Final Report. *Optica Acta*, **6**, 1-26.
- Stockman A, MacLeod DIA & Johnson NE (1993). Spectral Sensitivities of the Human Cones. *Journal of the Optical Society of America A*, **10** (12), 2491-2521.
- Stockman A & Sharpe LT (1999). Cone Spectral Sensitivity and Color Matching. In *Colour Vision*. ed. Gegenfurtner KR & Sharpe LT, Cambridge University Press, Cambridge, UK, 53-88.
- Stockman A & Sharpe LT (2000). Spectral Sensitivities of the Middle- and Long-Wavelength Sensitive Cones Derived from Measurements in Observers of Known Genotype. *Vision Research*, **40**, 1711-1737.
- Stockman A, Sharpe LT & Fach CC (1999). The Spectral Sensitivity of the Human Short-Wavelength Cones. *Vision Research*, **39**, 2901-2927.
- Sun C & Wee WG (1983). Neighboring Gray Level Dependence Matrix for Texture Classification. *Computer Vision, Graphics, and Image Processing*, **23**, 341-352.
- Sun Q & Fairchild MD (2004). Image Quality Analysis for Visible Spectral Imaging Systems. *The Journal of Imaging Science and Technology*, **48** (3), 211-221.
- Sung L-P, Nadal ME, McKnight ME, Nguyen JV & Lin C-J (2000). Investigation of the Relationship Between Microstructure and Appearance Properties of Coating Materials. In *Proceedings of ICE 2000-Federation of Societies for Coatings Technology Meeting Technical Program. Federation of Societies for Coatings Technology (FSCT) Meeting, Chicago, USA*, 534-539.
- Sung L-P, Nadal ME, McKnight ME, Marx E & Laurenti B (2002). Optical Reflectance of Metallic Coatings: Effect Of Aluminum Flake Orientation. *Journal of Coatings Technology*, **74** (932), 55-63.
- Tamura H, Mori S & Yamawaki T (1978). Texture Features Corresponding to Visual Perception. *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-8** (6), 460-472.
- Teaney S, Pfaff G & Nitta K (1999). Pearlescent Pigments: The Next Generation. *Modern Paint and Coatings*, **89** (8), 16-21.
- Thurstone LL (1959). *The Measurement of Values*. The University of Chicago Press, Illinois, USA, 19-49.
- Torgerson WS (1967). *Theory and Methods of Scaling*. John Wiley & Sons Inc., New York, USA, 205-246.

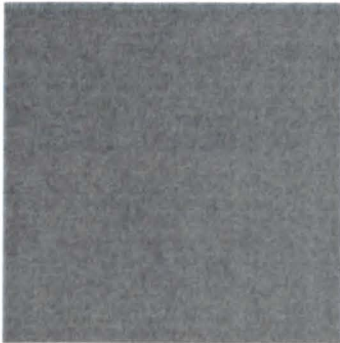
- Upton GJG & Cook I (1996). *Understanding Statistics*. Oxford University Press, Oxford, UK, 580-589.
- Valberg A (2005). *Light Vision Color*. John Wiley & Sons Ltd., West Sussex, UK, 105.
- van Aken H (2006). Color Measurement for the Coatings Industry. In *Coatings Technology Handbook*, 3rd Edition. ed. Tracton A, A. , CRC Press Inc., FL, USA, Section Ten; 1-2.
- van Meeteren A & Vos JJ (1972). Resolution and Contrast Sensitivity at Low Luminance Levels. *Vision Research*, **12**, 825-833.
- Venable WH (1987). A Model for Interpreting Three-angle Measurement of Flake Finishes. In *Proceedings of the ISCC Williamsburg Conference on Appearance*, 57-60.
- Virsu V & Rovamo J (1979). Visual Resolution, Contrast Sensitivity, and the Cortical Magnification Factor. *Experimental Brain Research*, **37**, 475-494.
- Walraven PL & Bouman MA (1966). Fluctuation theory of colour discrimination of normal trichromats. *Vision Research*, **6**, 567-586.
- Wandell BA (1995). *Foundations of Vision*. Sinauer Associates Inc., Massachusetts, USA, 29, 46, 145-150, 155-176, 195-246.
- Watanabe A, Sakara H & Isono H (1976). Chromatic Spatial Sine-wave Response of the Human Visual System. *NHK Lab. Note*, **198**, 1-10.
- Wei J (2006). Assessing the Appearance of Objects. *PhD Thesis*, University of Leeds, UK.
- Wen S & Wu R (2006). Two-Primary Crosstalk Model for Characterizing Liquid Crystal Displays. *Color Research and Application*, **31** (2), 102-108.
- Westland S (2002). Models for Colour Perception and Colour Appearance. In *Colour in Foods - Improving Quality*. ed. MacDougall DB, Woodhead Publishing Ltd., Cambridge, UK, 65-73.
- Westland S (2005). Models of the Visual System and Their Application to Image-Quality. In *Proceedings of AIC Colour 2005, the 10th Congress of the International Colour Association*. Granada, Spain, 309-312.
- Westland S & Ripamonti C (2004). *Computational Colour Science using MATLAB*. John Wiley & Sons Ltd., West Sussex, UK.
- Westlund HB & Meyer GW (2001). Applying Appearance Standards to Light Reflection Models. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques, SIGGRAPH*, NY, USA, 501-510.
- Wezka JS, Dyer CR & Rosenfeld A (1976). A Comparative Study of Texture Measures for Terrain Classification. *IEEE Transactions on Systems, Man and Cybernetics*, **SMC-6** (4), 269-285.

- Wright WD (1928). A Re-determination of the Trichromatic Coefficients of the Spectral Colours. *Transactions of the Optical Society*, **30**, 141-164.
- Wyszecki G & Stiles WS (2002). *Color Science: Concepts and Methods, Quantitative Data and Formulae*. 2nd Edition. John Wiley & Sons Inc, New York, USA, 83-91,101-102,.
- Xiao F, DiCarlo JM, Catrysse PB & Wandell BA (2002). High Dynamic Range Imaging of Natural Scenes. In *Proceedings of the 10th Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications*, Arizona, USA, 337-342.
- Xin JH, Shen H-L & Lam CC (2005). Investigation of Texture Effect on Visual Colour Difference Evaluation. *Color Research and Application*, **30** (5), 341-347.
- Yoshida Y & Yamamoto Y (2001). Proposal for Color Management of LCD. In *Proceedings of the 9th Color Imaging Conference: Color Science and Engineering: Systems, Technologies, Applications*, Arizona, USA, 233-238.
- Zhang X & Wandell BA (1997). A Spatial Extension of CIELAB for Digital Color Image Reproduction. *Journal of the Society for Information Display*, **5** (1), 61-63.

Appendices

Appendix I: Images of Metallic-Coating Panels used for Coarseness Assessments

The images of the metallic-coating panels used for the coarseness assessment. One image from each colour group (grey, blue, green and purple) is shown. It should be noted that the images may look different from the original metallic-coating panels.



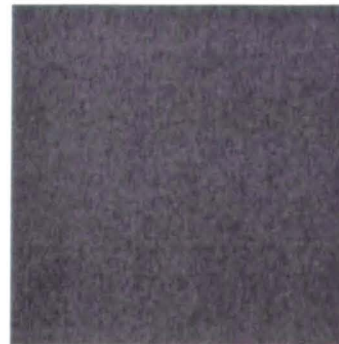
Grey Colour Sample
(No. 6)



Blue Colour Sample
(No. 6)



Purple Colour Sample
(No. 3)



Purple Colour Sample
(No. 3)

Appendix II: Scale Values and Model Predictions from Coarseness Assessments

The scale values for each sample obtained from the visual assessments of the coarseness using the metallic-coating panels (physical samples) and the image samples at Condition 6 and Condition 1 together with the model predictions of the corresponding samples and experimental conditions.

Sample	Perceptual Coarseness			Model Prediction		
	Physical Sample	Condition 6	Condition 1	Physical Sample	Condition 6	Condition 1
Grey Colour Sample						
1	3.15	3.78	4.51	0.381	0.419	0.315
2	3.60	4.04	4.62	0.397	0.436	0.329
3	5.75	5.91	5.80	0.505	0.524	0.434
4	7.10	6.83	6.06	0.574	0.582	0.503
5	7.70	7.74	6.93	0.659	0.660	0.589
Blue Colour Sample						
1	5.90	6.11	6.10	0.505	0.533	0.436
2	5.70	5.35		0.458	0.484	
3	7.30	6.98		0.596	0.615	
4	5.70	5.89	5.50	0.468	0.491	0.399
5	5.95	5.91		0.507	0.530	
6	7.65	7.41		0.637	0.656	
7	5.90	5.59		0.501	0.525	
8	7.50	7.09		0.642	0.653	
9	5.80	5.84		0.530	0.559	
10	7.45	6.98	6.91	0.631	0.651	0.562
11	5.60	5.45	5.55	0.494	0.527	0.426
12	6.75	6.82		0.608	0.626	
13	7.00	6.50	6.32	0.592	0.613	0.521
14	4.90	5.51		0.481	0.526	
15	6.30	6.21		0.530	0.555	
16	5.55	5.47	5.36	0.496	0.535	0.430
17	4.75	4.95		0.487	0.526	
18	6.40	6.46		0.535	0.564	
19	6.05	5.89		0.506	0.524	
20	5.95	5.88	5.80	0.483	0.509	0.412
21	5.65	5.77		0.480	0.494	
22	5.65	5.85		0.490	0.516	
23	5.65	6.11		0.504	0.534	
24	5.90	5.66	5.46	0.471	0.485	0.399
25	5.85	5.24	5.32	0.436	0.450	0.361
26	5.60	5.65	5.33	0.469	0.501	0.397

Sample	Perceptual Coarseness			Model Prediction		
	Physical Sample	Condition 6	Condition 1	Physical Sample	Condition 6	Condition 1
27	5.30	4.79	5.04	0.376	0.396	0.305
28	6.75	6.64		0.566	0.578	
29	5.55	5.47		0.421	0.439	
30	6.15	6.25	6.16	0.473	0.480	0.401
31	6.90	7.04		0.597	0.607	
32	5.45	6.14		0.486	0.513	
33	6.90	7.19		0.607	0.631	
34	5.65	6.19		0.488	0.516	
35	7.15	7.20	7.37	0.612	0.624	0.541
36	5.85	5.94	6.04	0.498	0.522	0.429
37	7.05	7.23	6.99	0.592	0.605	0.521
38	6.75	7.06	6.80	0.577	0.604	0.508
39	4.70	5.59	5.29	0.454	0.487	0.383
40	6.05	6.40	6.17	0.509	0.537	0.440
41	5.25	6.00		0.477	0.513	
42	5.60	6.22		0.492	0.522	
43	5.70	5.55		0.472	0.501	
44	5.00	5.48		0.443	0.468	
45	5.80	5.60	5.61	0.457	0.473	0.387
46	5.45	5.90	5.68	0.467	0.492	0.399
47	5.40	5.43		0.460	0.488	
48	5.25	5.84		0.468	0.490	
49	5.45	5.47		0.444	0.466	
50	5.30	5.08		0.404	0.436	
Green Colour Sample						
1	6.45	6.50		0.604	0.616	
2	5.70	5.93	5.60	0.546	0.559	0.470
3	7.85	7.80	7.31	0.720	0.723	0.646
4	5.95	6.21	5.99	0.559	0.571	0.483
5	6.65	6.68	6.54	0.590	0.610	0.519
6	8.05	7.73		0.730	0.741	
7	6.85	7.09	6.61	0.633	0.649	0.561
8	7.85	7.77	7.45	0.729	0.740	0.657
9	6.70	6.35		0.601	0.628	
10	7.70	7.80	7.61	0.718	0.727	0.645
11	6.95	7.39	7.10	0.641	0.658	0.569
12	7.85	8.18		0.726	0.740	
13	7.80	7.40		0.704	0.729	
14	5.80	5.50		0.545	0.587	
15	6.85	6.96		0.616	0.631	
16	6.25	6.86	6.32	0.577	0.612	0.507
17	5.95	6.10		0.567	0.596	
18	7.05	6.49		0.636	0.652	
19	6.55	6.65		0.612	0.622	
20	6.15	6.82	6.50	0.611	0.628	0.539
21	6.30	6.93		0.605	0.614	

Sample	Perceptual Coarseness			Model Prediction		
	Physical Sample	Condition 6	Condition 1	Physical Sample	Condition 6	Condition 1
22	6.25	6.63	6.31	0.593	0.611	0.519
23	6.35	6.68		0.603	0.612	
24	6.05	6.47		0.576	0.592	
25	6.25	6.74	6.29	0.552	0.567	0.479
26	6.30	7.01		0.596	0.617	
27	5.85	5.76		0.505	0.527	
28	7.75	8.07		0.710	0.715	
29	5.90	6.16		0.526	0.550	
30	6.00	6.63		0.572	0.587	
31	7.80	8.14		0.719	0.720	
32	6.50	6.89		0.597	0.612	
33	8.20	8.21	7.75	0.758	0.767	0.684
34	6.70	6.96		0.627	0.646	
35	8.10	7.99	7.77	0.744	0.758	0.671
36	6.65	6.78		0.610	0.632	
37	7.95	7.75		0.735	0.742	
38	8.15	7.99		0.744	0.759	
39	6.85	6.87		0.624	0.645	
40	7.25	7.24	6.94	0.638	0.660	0.567
41	6.70	6.88	6.70	0.614	0.637	0.541
42	6.80	7.07		0.620	0.642	
43	6.80	6.82	6.73	0.610	0.621	0.535
44	6.30	6.69	6.33	0.568	0.585	0.496
45	6.20	6.74		0.582	0.595	
46	6.25	6.70	6.59	0.583	0.594	0.509
47	6.60	7.01		0.596	0.611	
48	6.45	7.23		0.610	0.633	
49	6.15	6.48	5.98	0.542	0.561	0.469
50	5.80	5.88		0.508	0.524	
Purple Colour Sample						
1	4.10	4.36		0.395	0.422	
2	4.40	4.40		0.383	0.409	
3	5.50	5.65	5.68	0.493	0.496	0.422
4	4.35	4.51	4.70	0.393	0.419	0.326
5	4.05	4.84		0.393	0.423	
6	5.30	5.61		0.447	0.472	
7	3.20	4.41		0.357	0.396	
8	4.25	5.66	5.50	0.438	0.470	0.367
9	3.05	4.00		0.341	0.385	
10	3.90	4.77		0.392	0.434	
11	3.00	3.57	3.86	0.313	0.372	0.239
12	3.50	4.84	5.26	0.402	0.431	0.324
13	3.15	4.14	4.20	0.368	0.419	0.303
14	2.30	2.72		0.295	0.371	
15	4.15	4.99		0.426	0.456	
16	2.40	3.88		0.316	0.381	

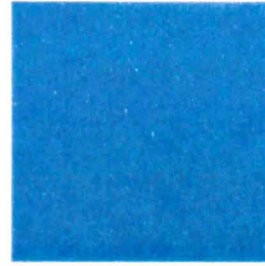
Sample	Perceptual Coarseness			Model Prediction		
	Physical Sample	Condition 6	Condition 1	Physical Sample	Condition 6	Condition 1
17	2.55	3.34		0.307	0.374	
18	4.80	4.71		0.413	0.447	
19	5.10	5.44	5.34	0.452	0.473	0.384
20	4.05	4.36	4.58	0.365	0.392	0.296
21	3.90	4.18		0.368	0.401	
22	3.40	4.00		0.365	0.400	
23	3.20	3.69	4.39	0.354	0.396	0.284
24	4.35	4.63	4.85	0.404	0.425	0.335
25	4.65	4.79	5.16	0.408	0.434	0.343
26	3.75	4.46	4.71	0.378	0.411	0.306
27	4.00	3.85		0.354	0.371	
28	5.25	5.71		0.453	0.454	
29	3.90	4.27		0.373	0.386	
30	3.80	4.35		0.384	0.405	
31	5.15	5.58		0.466	0.481	
32	3.65	4.52		0.355	0.387	
33	4.70	5.39		0.430	0.458	
34	3.30	4.18	4.54	0.346	0.381	0.270
35	4.00	5.23	5.29	0.411	0.443	0.342
36	3.20	4.45	4.79	0.360	0.391	0.285
37	3.50	4.94		0.379	0.422	
38	3.60	3.74		0.406	0.443	
39	2.50	2.75	3.39	0.294	0.364	0.218
40	3.35	4.83		0.379	0.416	
41	2.50	3.87	4.00	0.326	0.382	0.257
42	2.45	3.15		0.275	0.337	
43	4.30	4.52	4.76	0.416	0.437	0.342
44	4.45	4.42		0.393	0.413	
45	4.15	4.14		0.387	0.416	
46	4.30	3.99		0.393	0.413	
47	3.35	3.98	4.37	0.370	0.396	0.300
48	3.50	4.19		0.362	0.393	
49	4.60	4.37	4.29	0.385	0.412	0.317
50	4.90	4.57		0.391	0.405	

Appendix III: Images of Metallic-Coating Panels used for Glint Assessments

The images of the metallic-coating panels used for the glint assessment. One image from each colour group (grey, blue, brown green, red and yellow) is shown. It should be noted that the images may look different from the original metallic-coating panels.



Grey Colour Sample
(No. 6)



Blue Colour Sample
(No. 4)



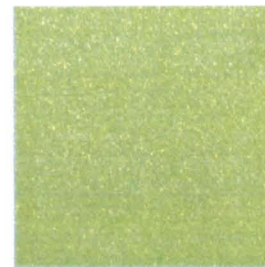
Brown Colour Sample
(No. 4)



Green Colour Sample
(No. 4)



Red Colour Sample
(No. 4)



Yellow Colour Sample
(No. 4)

Appendix IV: Scale Values and Model Predictions from Glint Assessments

The scale values for each sample obtained from the visual assessments of the perceptual glint using the metallic-coating panels and the model predictions of the corresponding samples. The model predictions given below were computed from the models incorporated with the iterative method with the metric *M6* and the linearisation using *Power2* at the threshold 56 based on the samples whose luminance values were obtained using the linear regression characterisation model and that at the threshold 57 based on the samples whose luminance values were obtained using the polynomial characterisation model; and from the models incorporated with the distribution-estimate method with the metric *M2* and *Power2* based on the samples whose luminance values were obtained using the linear regression characterisation model and with the metric *M2* and *Power1* based on the samples whose luminance values were obtained using the polynomial characterisation model.

Perceptual Glint		Model Prediction			
Segmentation Method		Iterative		Distribution-estimate	
Characterisation Model		Linear	Polynomial	Linear	Polynomial
Grey Colour Sample					
1	16.0	17.7	16.0	24.2	30.2
2	33.9	30.7	28.2	37.1	43.1
3	73.1	71.1	68.4	63.2	70.9
4	101.7	104.4	102.9	83.8	92.2
5	146.6	166.4	157.2	113.1	107.1
Blue Colour Sample					
1	8.9	11.9	12.5	20.1	31.2
2	102.4	103.0	101.2	105.9	101.6
3	20.7	40.5	38.9	50.9	39.6
4	144.2	139.1	158.9	103.3	145.8
5	50.7	55.0	58.1	60.7	80.5
6	73.1	76.1	75.9	69.3	76.9
7	31.8	39.1	35.8	40.7	39.7
8	58.1	67.8	69.1	63.7	79.8
9	130.5	134.0	128.8	100.7	102.9
10	2.8	-3.4	-2.3	17.5	22.6
11	131.2	152.6	145.9	122.2	113.5
12	91.5	91.6	92.6	69.9	80.1
13	14.2	21.4	21.3	28.0	38.1

Perceptual Glint		Model Prediction			
Segmentation Method		Iterative		Distribution-estimate	
Characterisation Model		Linear	Polynomial	Linear	Polynomial
14	137.7	144.6	138.0	126.6	118.0
15	5.2	3.6	4.1	15.4	23.8
16	96.9	97.7	97.8	79.7	85.9
17	4.3	7.8	8.7	12.9	23.6
18	47.7	52.5	51.9	47.8	58.7
19	41.2	37.1	33.9	38.9	38.7
20	32.6	29.4	28.7	44.1	50.4
Brown Colour Sample					
1	6.2	4.0	4.6	6.0	9.1
2	87.5	82.8	83.9	85.8	83.0
3	12.9	21.6	22.0	10.7	18.2
4	116.8	98.5	107.5	114.4	125.9
5	43.2	44.8	38.9	62.6	33.6
6	62.5	61.7	58.6	76.0	58.3
7	14.7	19.8	24.6	15.0	50.8
8	48.8	52.7	54.7	32.5	44.5
9	110.0	117.6	117.7	145.8	126.4
10	2.6	0.3	1.6	15.1	22.2
11	119.2	120.0	117.1	144.5	118.3
12	78.2	72.6	73.3	94.7	90.0
13	8.5	3.8	4.9	20.0	25.8
14	123.7	111.1	118.8	131.6	135.3
15	4.2	11.9	12.8	7.6	16.2
16	80.1	61.3	62.2	69.6	71.5
17	4.3	11.6	11.7	7.9	12.5
18	41.6	35.6	31.0	57.4	36.9
19	32.4	29.3	29.8	41.1	47.2
20	22.6	28.8	30.9	27.9	42.8
Green Colour Sample					
1	5.1	11.5	11.8	20.4	29.9
2	84.7	89.3	84.5	90.8	76.2
3	15.9	18.2	18.6	19.1	29.5
4	133.4	130.4	144.5	121.4	147.4
5	48.0	53.3	52.0	54.7	57.9
6	71.2	70.3	76.8	72.3	98.1
7	23.9	43.6	45.8	42.4	62.3
8	66.0	69.0	73.4	66.5	85.1
9	163.2	177.1	181.9	180.8	168.1
10	2.8	-0.3	0.6	11.4	15.5
11	135.3	160.5	147.2	193.9	140.8
12	74.8	76.4	82.0	62.8	84.4
13	8.8	13.1	11.4	24.5	22.6
14	141.9	149.2	146.0	145.0	128.6
15	4.2	11.4	10.7	21.9	17.2

Perceptual Glint		Model Prediction			
Segmentation Method		Iterative		Distribution-estimate	
Characterisation Model		Linear	Polynomial	Linear	Polynomial
16	91.2	98.7	91.9	113.7	90.5
17	5.0	12.4	11.8	16.0	15.5
18	47.0	57.3	54.5	68.4	62.5
19	33.0	44.3	41.8	63.3	61.4
20	23.7	31.0	31.9	45.3	60.4
Red Colour Sample					
1	4.9	-1.9	-0.8	6.1	13.8
2	82.3	62.4	60.6	54.2	57.0
3	13.3	7.9	7.0	10.4	21.4
4	163.9	128.4	126.0	113.2	102.8
5	36.0	15.3	15.3	24.2	32.4
6	60.3	45.9	42.2	43.0	38.5
7	19.0	15.0	15.4	20.2	30.3
8	46.4	45.6	43.7	39.7	38.9
9	107.5	95.6	92.0	66.9	65.2
10	2.6	5.0	4.6	12.8	14.8
11	110.2	75.1	72.8	60.3	65.1
12	73.8	47.9	50.5	35.1	54.6
13	7.6	0.4	1.4	7.6	17.8
14	118.7	103.6	101.3	77.9	77.2
15	3.2	0.4	1.6	5.2	16.1
16	88.7	72.5	75.2	58.1	76.9
17	5.4	4.7	5.2	5.4	9.6
18	38.8	18.9	16.1	22.6	27.4
19	23.6	14.2	13.4	21.1	28.2
20	21.4	5.2	5.1	16.8	24.1
Yellow Colour Sample					
1	5.1	11.2	11.7	4.2	10.0
2	75.8	83.6	85.3	89.1	87.0
3	8.8	18.9	22.2	15.9	41.4
4	115.4	96.8	98.8	82.7	81.1
5	35.7	48.4	48.4	45.7	48.2
6	57.5	75.9	77.0	30.0	35.6
7	14.5	30.4	30.9	26.2	34.3
8	48.8	57.8	58.2	55.2	57.7
9	141.6	128.7	133.8	146.2	133.4
10	2.3	9.0	9.3	15.3	22.4
11	139.2	115.5	123.6	132.5	133.2
12	67.9	72.4	71.2	36.2	39.9
13	7.0	5.6	4.9	4.9	7.5
14	115.9	93.4	90.4	104.4	85.7
15	2.8	15.6	16.3	5.1	12.1
16	77.9	62.5	63.6	54.1	57.5
17	3.4	6.9	7.7	1.7	8.6

Perceptual Glint		Model Prediction			
Segmentation Method		Iterative		Distribution-estimate	
Characterisation Model		Linear	Polynomial	Linear	Polynomial
18	28.3	44.9	43.2	52.2	46.5
19	25.7	47.2	46.3	51.1	49.4
20	18.5	31.7	32.4	24.7	32.1