Computational Model for Perceptual Coarseness Prediction

*S. Kitaguchi, M.R. Luo, *E.J.J. Kirchner and *G.J. van den Kieboom*
Department of Colour and Polymer Chemistry, University of Leeds (UK)
**Technology Center Colorimetry, Akzo Nobel Coatings, Car Refinishes (The Netherlands)*

Abstract

We present a computational model to predict perceptual coarseness from an image. The model was based on the hypothesis that a model for predicting perceptual coarseness should be motivated by human visual system. We found that the amplitude of the Fourier transform of an image captures information of coarseness. Furthermore, an analysis of Fourier amplitudes in terms of the human contrast-sensitivity function (CSF) leads to a metric that can predict perceptual coarseness. Model performance was proved by comparing the perceptual coarseness that was predicted by the model from images of metallic paint panels and the psychophysical data which was obtained by a visual assessment using the physical panels.

Introduction

Image texture analysis has been studied over the last few decades. Many image-texture analysis methods in the computing domain are based on various statistical analyses of the image itself with little attention given to the perception of texture [1]. However, psychophysical research demonstrates that the human visual system processes images in a way that is consistent with a spatial-frequency analysis of an image [2]. The contrast-sensitivity function (CSF) is a well established characteristic of the human visual system which describes the relationship between spatial frequency and contrast [3]. The contrast sensitivity threshold is the lowest contrast detectable at a given spatial frequency, and CSF defines differences in contrast sensitivity (sensitivity being inversely related to threshold) as a function of spatial frequency. It also has been used for analyzing perceptual differences in quality; colour difference and sharpness [4,5,6]. The square-root integral (SQRI) metric has been proposed by Barten [7] to evaluate the effect of the resolution on perceived image quality taking into account the CSF as shown in Equation (1).

$$SQRI = \frac{1}{\ln 2} \int_{0}^{u_{\max}} \sqrt{\frac{M(u)}{Mt(u)}} \frac{du}{u} \qquad (1)$$

where u is the spatial frequency in cycle/degree, u_{max} is the maximum spatial frequency to be displayed on a monitor, $M(u)$ is the modulation threshold function of the display and $Mt(u)$ is the modulation threshold function of the eye (CSF is the inverse of the modulation threshold function of the eye). This method can describe perceptual image differences according to changes in viewing distance and the average luminance of an image, *etc*. However, SQRI is independent of the actual image spatial-frequency data.

The objective of this study is to develop a computational model to predict perceptual texture from a digital image. As a target attribute for this model, coarseness was selected which was one of the essential terms for perceptual texture. The model should take in account the human visual system to analyse an image and should make a quantitative match with psychophysical data.

The coarseness model was based on the hypothesis that the amplitude in the Fourier transform of an image is a measure of the amount of contrast in the image and that the amount of contrast is correlated closely with perceptual coarseness. The model also assumed that CSF can be used to appropriately weight the importance of the contrast at the each spatial frequency. In order to test the model, a psychophysical experiment was carried out to scale the perceptual coarseness of metallic paint panels as samples.

Psychophysical Data & Analysis

Psychophysical data to be fitted for a coarseness model were obtained by visual assessment. A total of 10 observers (4 female and 6 male) with the normal colour participated in the experiment to scale perceptual coarseness. Metallic paint panels, having various coarseness levels and colours, were used as samples. Totally, there were 156 panels including 6 grey, 50 purple, 50 green and 50 blue panels. A DigiEye® viewing cabinet was used for this experiment, which incorporates diffuse light from two light sources (CIE illuminant D65 simulators) covered by diffusing filters and a flat base to present samples as shown in Figure 1 [8]. The reason of using the diffused light is to avoid any specular reflection or gloss which could distract observers during the coarseness assessment. The samples were presented on the bottom of the cabinet. An observer looked down onto the samples from the viewing window. The distance from the observer's eye to the sample was about 54cm. Categorical judgment [9] was applied to scale the perceived coarseness. Two metallic paint panels were presented for each trial in the viewing cabinet. One was a reference sample and the other was a test sample. One of 6 grey colour panels which had a middle coarseness level was used as a reference sample. Observer was asked to assign a category for a test sample comparing with the reference sample, whose category was 5, according to the observer's perception in terms of coarseness on a 1-9 scale as shown in Table 1. All samples were presented in a random order. To check the repeatability, each observer carried out the assessment twice. A total of 3100 (10 observers × 2 sessions × 155 samples) categorical judgments were made. Repeatability (intra-observer agreement) was investigated by calculating coefficient of determination (R-squared value) between the results of each observer's first session and second session and was found to be 0.69. Observer accuracy (inter-observer agreement) between each observer's data and the average of all observers' data was found to be 0.82 in R-squared values. The arithmetic mean of each observer's data was used as a measure of the perceptual coarseness for each sample.

Table 1: 1-9 categories used for the visual assessment

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

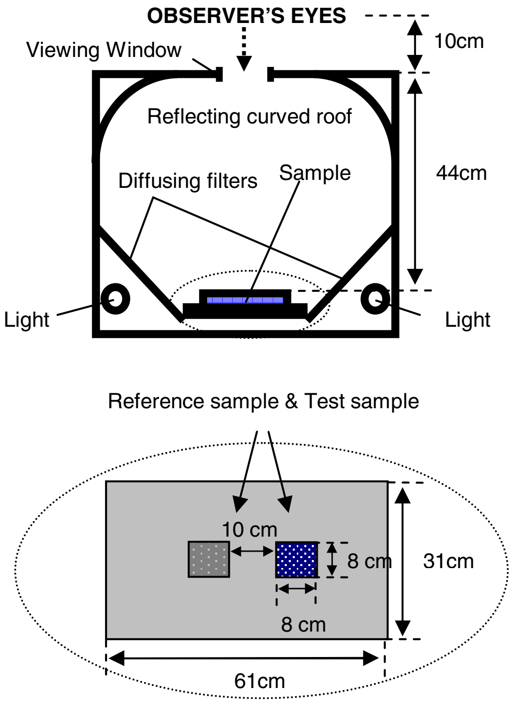

Figure 1. Schematic diagram of a DigiEye® Viewing Cabinet (top). Samples are placed on the base. Two light sources are positioned at the 2 bottom corners at each side and both emit light to the walls. The top corners have curved surfaces to reflect light uniformly onto the sample. An observer looked down the sample from the viewing window. An illustration of samples' arrangement and an observer's viewing field (bottom).

Model Design

We outline our model in this section. First the necessary data for the model are given. Successively, each stage of the model is introduced.

Test Images & Data

A model was developed to predict the perceptual coarseness from an image of an object captured by a digital camera. Images of the samples (metallic paint panels) used for the visual assessment were captured in the DigiEye® viewing cabinet. Therefore, the capturing condition was consistent with the visual assessment. The camera was located at the viewing window where the observer's eyes were as illustrated in Figure 1. A tele-spectroradiometer (TSR) was used to measure the tristimulus values for each sample. It was also located at the viewing window of the DigiEye® viewing cabinet. In this experiment, the DigiEye® viewing cabinet was always used all the visual assessment, the image capturing and the measurement by the TSR, because any viewing geometry difference could seriously affect the appearance of the metallic paint panels. In this work, XYZ values of each pixel of each sample image were also obtained by transforming the camera RGB values using a camera characterisation model*, i.e.* a polynomial model using the least squares method [10]. An example of a linear mapping method between camera RGB values and XYZ values is shown as Equation (2). This linear transformation is a special case of the set of polynomial transforms. Practically, higher order (non-linear) transformations were used.

$$\begin{aligned} X &= a_{11}R + a_{12}G + a_{13}B \\ Y &= a_{21}R + a_{22}G + a_{23}B \\ Z &= a_{31}R + a_{32}G + a_{33}B \end{aligned} \tag{2}$$

where the coefficients a_{11}~ a_{33} were empirically determined.

Usually, a standard chart such as GretagMacbeth Color Checker Digital Chart or a set of Munsell colours is often used as training data to determine the coefficients. However, in this study, the average RGB values of the sample images and the corresponding measured XYZ values of the physical samples (the metallic paint panels) were used as a training data set to derive the model. Since surface material differences could affect the performance of the camera characterisation model, a model derived from a chart which has a matt surface may not be applicable for other samples having a glossy surface such metallic paints.

The input parameters needed for a computational coarseness model were the sample size (8×8 cm^2), the viewing distance (54 cm) and the luminance of the reference white tile (167.8 cd/m^2) corresponding to the viewing conditions in the visual assessments.

Channel Selection

After the transform from RGB values to XYZ values was completed, the XYZ values were encoded into the cone-excitation space LMS corresponding to the long-(L), middle-(M) and short-(S) wavelength responses in the cone spectral sensitivity of human visual system, using the Stockman, MacLeod and Johnson transformation [11]. The image was then separated into three channels, *i.e.* the luminance channel, the red-green channel and the yellow-blue channel according to a chromaticity coordinate system proposed by MacLeod and Boynton [12] as shown in below.

Luminance Channel = L+M

Red – Green Channel = L / (L+M)

Yellow - Blue Channel = S / (L+M)

Before applying the Fourier transform to the image, the mean value for each channel was subtracted from every pixel value in each channel respectively, in order that the DC component in the Fourier transform was zero. A two-dimensional discrete Fourier transform (DFT) was applied to each of the three channels (luminance, red-green and yellow-blue) to transfer the spatial domain into the frequency domain. Figure 2 shows an original image of a sample and its Fourier spectrum images for the three channels. These images indicate that there is little Fourier energy in either chromatic channels and that there is huge amount of Fourier energy in the luminance channel. This suggests that, for our samples, the chromatic channels have little contribution to perceptual coarseness. Therefore, this study focuses only on the luminance channel for modelling coarseness.

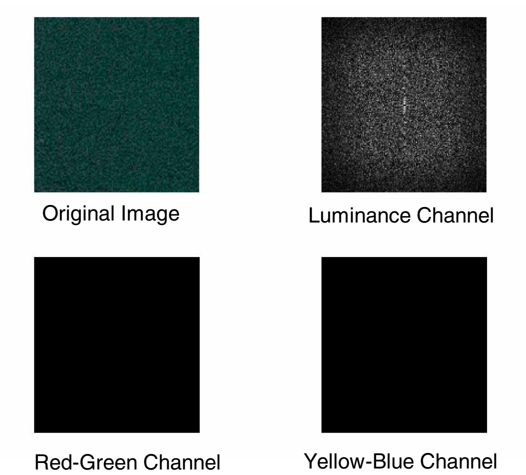

Figure 2. An original image of a metallic paint panel (top left) and its Fourier spectrum images for luminance (top right) and chromatic channels: red-green channel (bottom left) and yellow-blue channel (bottom right). Note that the DC component is in the centre of each Fourier spectrum image and that spatial frequency increase from the centre to outwards.

Applying CSF

To incorporate properties of the human visual system into the model, the Fourier energy was weighted using the CSF [13] as shown in Equation (3), and the sum of these weighted values was computed.

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

$$\begin{aligned}
\text{where }\; & F_L = 1000(L/70)^{1/3} && \text{if } 1 \le L \le 70 \\
& F_L = 1000(1/70)^{1/3} && \text{if } L < 1 \\
& F_L = 1000 && \text{if } L > 70 \\
& F_C = (1-d)^2 \\
& d = [\,(x - x_{white})^2 + (y - y_{white})^2\,]^{0.5}
\end{aligned} \qquad (3)$$

where u is the spatial frequency in cycle/degree, L is the luminance of the stimulus in units of cd/m^2, d measures how chromatic the image is, (x, y) is the average chromaticity co-ordinate derived from the XYZ values for an sample image and (x_{white}, y_{white}) is the white point.

Finally, the sum of the weighted values was normalised using the mean value of the luminance channel according to the human contrast sensitivity, in which the ratio of the increment threshold to the background intensity is said to be a constant. This can be illustrated by $\Delta P/P = K$ where ΔP represents the difference of threshold; P is intensity of the background; K is a constant. This equation is in the form same as Weber's Law [2].

This phenomenon can be clearly recognized by comparing the Fourier energy weighted by the CSF and the mean value of the luminance channel as shown in Figure 3. The 4 points in Figure 3 correspond to the 4 sample images, which have the perceptual coarseness values of 5.15-5.25 from the psychophysical experiment (in a 1-9 scale). According to the psychophysical data, these images are expected to be similar in appearance. However, it is evidenced that, as shown in Figure 3, the Fourier energy weighted by the CSF is much greater for the bright sample images. Therefore, this effect was included in the model by normalising the Fourier energy using the mean value of luminance channel for each image. It can be expressed by Equation (4).

$$NM = \sum_{0}^{u_{max}} \frac{E(u) \times CSF(u)}{I \times S} \qquad (4)$$

where u is the spatial frequency in cycle/degree, u_{max} is the maximum spatial frequency containing in an image, $CSF(u)$ is the CSF given in Equation (3), $E(u)$ is the Fourier energy, I is the mean value of the luminance channel and S is the size of an image in pixel units.

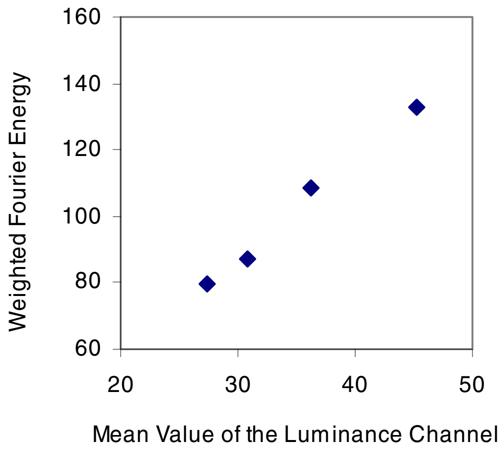

Figure 3. An example of the relationship between Fourier energy and the brightness of sample images having similar perceptual coarseness (5.15 – 5.25 scale values obtained from the visual assessment).

Model Performance

The performance of the model was evaluated by comparing the perceptual coarseness that was predicted by the model (NM) from the sample images with the psychophysical data from the visual assessing using physical samples (Figure 4). It is notes that the NM model predictions in Figure 4 were normalised using the maximum. The comparison showed a non-linear relationship between the model prediction and the psychophysical data. Therefore, the model was extended to incorporate a final stage of linearisation.

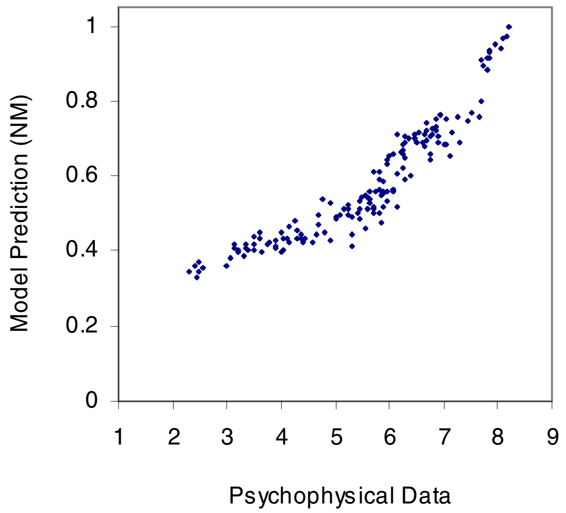

Figure 4. Comparison between the predicted coarseness by the model (NM) and the visual result for all samples.

Linearisation

The model (NM) was derived based on the spatial frequency information and the characteristics of the human visual system. At this stage, it tried to establish a linear relationship between the previous model's (NM) prediction and the psychophysical data. This was achieved by taking the logarithm of the model (NM) resulting in the linearised coarseness model (CM) given in Equation (5).

$$CM = \log(NM) \tag{5}$$

The performance of linearisation was given in Figure 5 (the CM model predictions were normalised using the maximum). An R-squared value of 0.91 for all the samples indicates the excellent relationship between the model prediction and the psychophysical data. In Figure 6, the performances of the linearised model (CM) for grey, purple, green and blue samples are given individually. The R-squared values are 0.96, 0.80, 0.95 and 0.79 for the grey, purple, green and blue sample respectively. The grey samples result shows the best performance. It is because that there were only 6 grey samples and also they have clearly large perceived coarseness differences comparing with the other colours. Overall, the model shows good predictions for not only the grey samples but also for the coloured samples, even though only the luminance channel of the images was used.

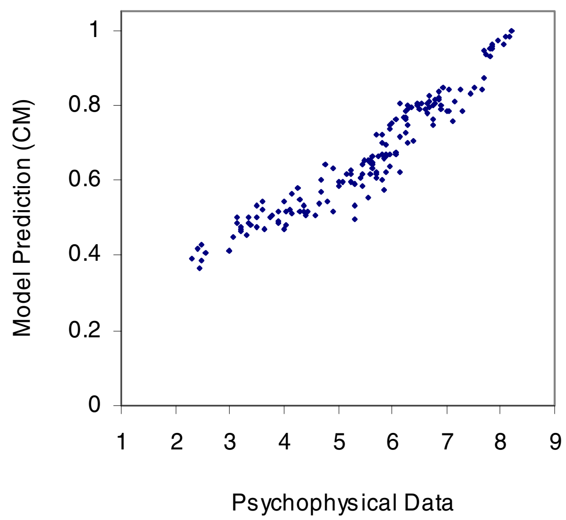

Figure 5. Comparison between the linearised model (CM) prediction and the visual result for all samples.

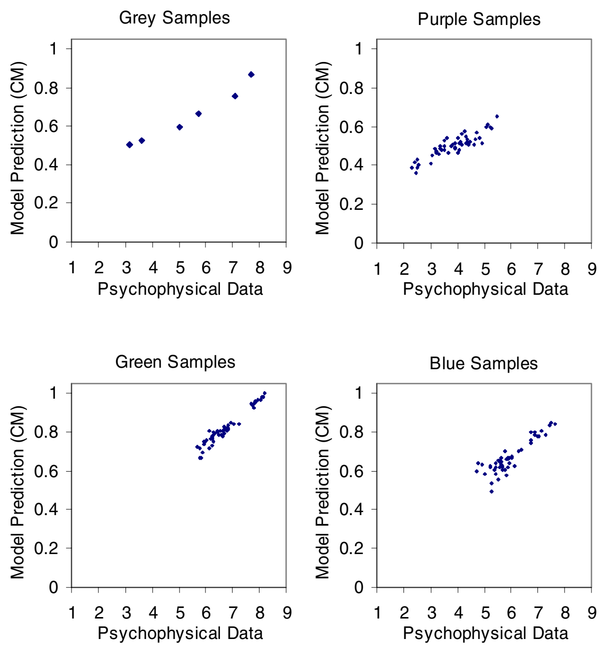

Figure 6. Comparison between the linearised model (CM) prediction and the visual result for grey, purple, green and blue samples individually.

Conclusion

In this paper, a computational model for predicting the perceptual coarseness was developed which not only analyses images, but also takes into account the human visual system which hasn't been included in most conventional image texture analysis methods. Figure 5 proves that the model can predict the perceptual coarseness of coloured samples as well as grey samples. It indicates that the luminance channel alone is sufficient for prediction coarseness. While the performance of the model was proved to be robust for metallic paint panels, further work are necessary to test the generality of the model using various types of texture and materials, and also the model is needed to be investigate the ability for changing in viewing distance which affect to appearance of texture.

Acknowledgements

This work was supported by Akzo Nobel. The authors would like to thank Professor Stephen Westland for his helpful comments.

References

[1] Tuceyran M. and Jain A.K., “Texture analysis”, in The Handbook of Pattern Recognition and Computer Vision (2nd Eds.), C.H. Chen, L.F. Pau and P.S.P. Wang, chapter 2, pp.235-276, World Scientific Publishing Co., 1998.

[2] Wandell B., “Foundations of Vision”, Sinauer Associates Inc, Sunderland, MA, 1995.

[3] Campbell F.W. and Robson J.G., Applications of Fourier Analysis to the visibility of gratings, J. Physiol. (Lond.), 181, pp.576-593, 1968.

[4] Bouzit S. and MacDonald L.W., “Sharpness Enhancement trough Spatial Frequency Decomposition”, IS&T’s PICS Conference, April 2001.

[5] Sun Q. and Fairchild M.D., “Image Quality Analysis for Visible Spectral Imaging Systems”, Journal IS&T, 48, 3, pp. 211-221, May/Jun 2004.

[6] Zhang X. and Wandell B.A., “A spatial extension of CIELAB for digital color reproduction”, proc. Of the SID symposiums, 1996.

[7] Barten P.G.J. , “Evaluation of subjective image quality with the square-root integral metjods”, J. Opt. Soc. Am. A, 7, 10, October 1990.

[8] Luo M.R., Cui G.H. and Li C., “British Patent: Apparatus and method for measuring colour (DigiEye system)”, Derby University Enterprise Limited, October 2001.

[9] Engeldrum P. G., “Psychophysical Scaling: a toolkit for imaging systems”, Optics & Photonics News, Imcotek Press; USA, 2000.

[10] Cheung V., Westland S., Connah D. and Ripamonti C., “A comparative study of the characterisation of colour cameras by means of neural networks and polynomial transforms”. Color.Technol. 120, pp.19-25, 2004.

[11] Stockman A. and MacLeod D.I.A. and Johnson N.E., “Spectral sensitivities of the human cones”, J. Opt. Soc. Am. A, 10, 12, pp.2491-2521, December 1993.

[12] MacLeod D.I.A and Boynton R.M., “Chromaticity diagram showing cone excitation by stimuli of equal luminance”, J. Opt. Soc. Am. , 69, 8, pp.1183-1186, 1979.

[13] Westland S., “Models of the visual system and their application to image-quality assessment”, Proceedings of the 10th Congress of the International Color Association, pp.309-312, Granada, Spain, 2005.

Author Biography

Saori Kitaguchi is a PhD student in Colour and Polymer Chemistry Department, University of Leeds, UK. She has a BSc degree in Chemistry and Materials Technology from Kyoto Institute of Technology in Japan and an MSc degree in Colour & Imaging Science from Colour & Imaging Institute, University of Derby, UK.