Adding Texture to Color: Quantitative Analysis of Color Emotions

Lucassen, M.P.; Gevers, T.; Gijsenij, A.

Publication date
2010

Document Version
Final published version

Published in
Final programme and proceedings: 5th European Conference on Color in Graphics, Imaging, and Vision, 12th International Symposium on Multispectral Color Science: June 14-17, 2010, Joensuu, Finaland, University of Eastern Finland

Adding texture to color: quantitative analysis of color emotions

Marcel P. Lucassen, Theo Gevers, Arjan Gijsenij; Intelligent System Laboratory Amsterdam, University of Amsterdam; Amsterdam, The Netherlands

Abstract

What happens to color emotion responses when texture is added to color samples? To quantify this we performed an experiment in which subjects ordered samples (displayed on a computer monitor) along four scales: Warm-Cool, Masculine-Feminine, Hard-Soft and Heavy-Light. Three sample types were used: uniform color, grayscale textures and color textures. Ten subjects arranged 315 samples (105 per sample type) along each of the four scales. After one week, they repeated the full experiment. The effect of adding texture to color samples is that color remains dominant for the Warm-Cool, Heavy-Light and Masculine-Feminine scale (in order of descending dominance), the importance of texture increases in that same order. The Hard-Soft scale is fully dominated by texture. The average intra-observer variability (between the first and second measurement) was 0.73, 0.66 and 0.65 for the uniform color, grayscale texture and color texture samples, respectively. The average inter-observer variability (between an observer and the other observers) was 0.68, 0.77 and 0.65, respectively. Using some 25,000 observer responses, we derived analytical functions for each sample type and emotion scale (except for the Warm-Cool scale on grayscale textures). These functions predict the group-averaged scale responses from the samples' color and texture parameters. For uniform color samples, the accuracy of our functions is significantly higher (average adjusted R^2 = 0.88) than that of functions previously reported. For color texture, the average adjusted R^2=0.80.

Introduction

There is growing interest in the understanding of human affective feelings in response to seeing colors. The so called 'color emotions' (i.e. emotional responses to color) involved in published studies do usually not refer to basic human emotions like happiness, surprise or fear. Rather, they capture the response on an associated affective dimension specified by the investigators and may therefore be anything. Color emotion studies recently published (e.g. [1]-[3]) have focused on the selection of emotional scales and how they interrelate (by means of factor analysis). Regression analysis shows the relationships of these scales with the underlying color appearance attributes (lightness, chroma and hue). Additionally the question whether color emotions can be regarded as culture specific or universal has been studied [1], [3]-[6]. Roughly summarizing the published studies on color emotion, the common finding is that the color emotions are reasonably well described by a small number of semantic factors, like for instance colour weight, colour activity and colour heat in [1]. Of the perceptual attributes, *lightness* and *chroma* are most frequently reported as being the relevant parameters for quantitative prediction of the color emotions, although *hue* cannot be ignored in scales like warm-cool.

So far, the role of texture in color emotion has received only little attention. An early study by Tinker [7] showed that surface texture, as represented by coated paper or cloth, had little or no effect upon apparent warmth or affective value of colors. Kim et al. [8] used color and texture features to predict human emotions based on textile images. Erhart & Irtel [9] indicate that surface structure can change the emotional effect of colored textile samples, depending upon the color. More recently, Simmons & Russell [10] reported that the addition of texture can significantly change the perceived unpleasantness of colors, depending on the texture class.

This paper reports on our study into the effect of texture on color emotion. Our systematic experiments build upon-, but differ in a number of ways from previous studies. Most importantly, instead of only studying uniformly colored samples, we also use samples with grayscale textures and color textures. These textures were synthesized to prevent biased responses such as reported in [10]. Second, we introduce a method in which all samples (shown on a computer display) maintain visible during experimental trials so that they can be ordered conveniently along an emotion scale. Third, our subjects performed the full experiment twice, with at least one week in between the first and second measurement. This allows quantification of the intra-observer variability over time, on which we are the first to report. We believe that repeatability information is at least as important as the information obtained from more observers. Finally, we sampled the available color gamut of our display in a very systematic manner to optimally cover both lightness, chroma and hue.

We analyze our data in terms of rank correlations within subjects and between subjects and provide quantitative descriptions. We derive color and texture emotion formulae that predict the group-averaged responses on the emotion scales from the samples' color and texture descriptors.

Methods

One of the problems we encountered in a pilot experiment is that when samples are shown one after the other, subjects tend to forget what responses they gave on the emotion scales for similar samples shown earlier in the trial. This leads to an unnecessary increase in variability in the subjects' responses, and therefore lower intra- and inter-observer correlations. To overcome this we designed our experiment in such a way that all samples maintain visible during a trial. We asked our subjects to order 105 square samples horizontally along an emotion scale labeled with opposite word pairs, using the computer mouse to drag samples from their initial location on the top of the screen. Samples could be dragged to any position on the screen to allow the subjects to keep an overview of the arrangement of the samples. Subjects knew that only the horizontal position of the samples on the scale would be analyzed.

Four emotion scales (Warm-Cool, Masculine-Feminine, Hard-Soft and Heavy-Light), in four different trials, were used. There were three conditions differing only in the type of samples. In the Uniform Color (UC) condition, uniformly colored samples were used that were systematically selected from the sRGB color gamut. In the Grayscale Texture (GT) condition the samples had a texture created in luminance, but not in the chromatic domain. Textures were generated using Perlin noise [11]. The third condition (Color Texture, CT) added a single color to the grayscale textures.

Examples of results obtained for these three conditions are shown in Figure 1.

Sample Selection

Uniform Color

The color monitor we used to display the samples was calibrated to the sRGB color space. Within the sRGB color gamut we selected 100 chromatic samples and 5 achromatic samples. Chromatic samples were selected at 5 lightness levels (L*=10,30,50,70,90). For each of these levels in L*, 10 hue angles were selected at 36 degree interval (h=0, 36, 72, …, 288, 324). Finally, for each of these hue angles two levels in C* were selected, being the maximum value within the sRGB gamut, denoted by C^*_{max}, and $C^*_{max}/2$. Five additional achromatic samples were selected at L*=20, 40, 60, 80, 100.

Grayscale and Color Texture

At this point in our research we did not want to use natural textures to avoid the possibility of strong inherent emotional associations. Therefore, we synthesized textures on the basis of Perlin noise [11]. Perlin noise is a primitive structure used in procedural texture generation, and is pseudo-random in appearance. All visual details in Perlin noise are the same size, which means that theoretically such an image can be said to truly represent a single texture. Perlin noise can be fully parameterized implying that we can reliably generate a random sample of textures by randomly sampling from the Perlin parameter space. Through controlling the number of octaves, the frequency of each octave and the amplitude of each octave we can respectively control the level of detail, the granularity and the contrast of the resulting texture. The Grayscale Textures were achromatic, having only spatial variations in lightness L*. Our Color Texture samples were colored versions of the Grayscale Textures (see Fig. 1).

Emotional Scale Selection

Contrary to previous studies, our primary aim is not to find out which scales are most appropriate to capture color emotions, but rather to explore the effects of adding texture to color samples. We therefore selected four scales with opposite word-pairs that have been frequently used in previous studies and for which we also gained experimental confidence in our pilot studies. These scales are Warm-Cool, Masculine-Feminine, Hard-Soft and Heavy-Light. The Warm-Cool scale was not used for the Grayscale Texture samples, because subjects found this combination very hard, if not impossible. With the exception of the scale Masculine-Feminine, quantitative descriptions of the scales on the basis of CIELAB parameters are available from previous studies, which enable us to compare results.

Subjects

Ten subjects participated in the experiments, 6 men and 4 women. Their ages ranged from 26 to 53, with an average of 31.9. Subjects were from 7 different nationalities: Dutch (4), Chinese (1), Russian (1), Italian (1), Spanish (1), Polish (1) and German (1). All subjects had normal color vision and normal or corrected to normal visual acuity. Subjects were screened for color vision deficiencies with the HRR pseudo-isochromatic plates (4^{th} edition), allowing color vision testing along both the red-green and yellow-blue axes of color space [12]. The HRR test was viewed under prescribed lighting (CIE illuminant C) using the True Daylight Illuminator (Richmond Products, Inc.), while illumination by other light sources was reduced to a minimum. The first author also participated as a subject in the experiment; the other subjects were unaware of the purposes of the experiment. Subjects participated on voluntary basis and did not receive a financial reward; they were all employed or studying at the institute where the experiment was carried out.

Monitor

Stimuli were presented on a high-resolution (1600x1200 pixels, 0.27 mm dot pitch) calibrated LCD monitor, an Eizo ColorEdge CG211. The monitor was driven by a computer system having a 24-bit (RGB) color graphics card operating at a 60 Hz refresh rate. Colorimetric calibration of the LCD was performed before each experimental session using the Eye-one spectrophotometer (GretagMacbeth). The monitor was calibrated to a D65 white point of 80 cd/m^2, with gamma 2.2 for each of the three color primaries. The CIE 1931 x,y chromaticities coordinates of the primaries were (x,y) = (0.638, 0.322) for red, (0.299, 0.611) for green and (0.145, 0.058) for blue, respectively. With these settings of our monitor we closely approximate the sRGB standard monitor profile [13]. Spatial uniformity of the display, measured relative to the center of the monitor, was $\Delta E^*_{ab} < 1.5$ according to the manufacturer's calibration certificates.

Procedure

Subjects were seated in front of the monitor at a viewing distance of about 60 cm. The screen size extended 39.6° x 30.2° visual angle, and a sample 2.6° x 2.6°. Samples were initially displayed in random order at the top of the screen. Subjects dragged the samples away from their initial position to give them a relative ordering along the horizontal emotion scale. Subjects knew that only the horizontal position would be analyzed, the vertical space could be used to keep an overview of the samples. After ordering the first group of 50 samples, subjects pressed a button that showed the second group of 55 samples (the first 50 samples remained visible). During a trial all samples could be re-ordered if desired. On average, one trial of 105 samples took about 5-10 minutes. All subjects repeated the experiment with at least one week in between the first and the second measurement.

Results

Examples of the results for a single observer are shown in Fig. 1. Actual scale values for the samples were calculated from their horizontal midpoints. Throughout this paper we use ranks (i.e. a relative order from the left side to the right side of the scale) and rank correlations rather than the absolute scale values, because the scales are not expected to be linear. An additional advantage of using ranks is that it corrects for individual differences in the used scale range. For instance, one subject may use the full scale range to position the samples, while another subject may use only 75% of that range. Statistical analyses were performed with the Statgraphics Centurion XV software package.

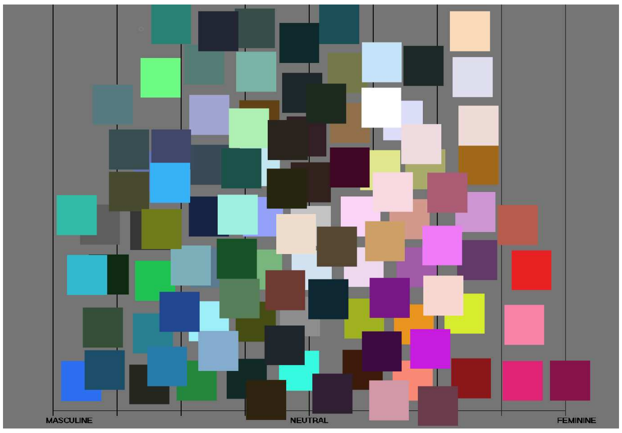

Figure 1a. *Experimental result (data from a single observer) for the Uniform Color samples, ordered horizontally along the Masculine – Feminine scale. Only the horizontal position matters. At a viewing distance of 60 cm, the screen size extended 39.6° x 30.2° visual angle, and one sample 2.6° x 2.6°. The 100 chromatic patches systematically sampled the sRGB color gamut at 5 lightness levels, 10 hue levels, and 2 chroma levels. Additionally, 5 achromatic samples were used.*

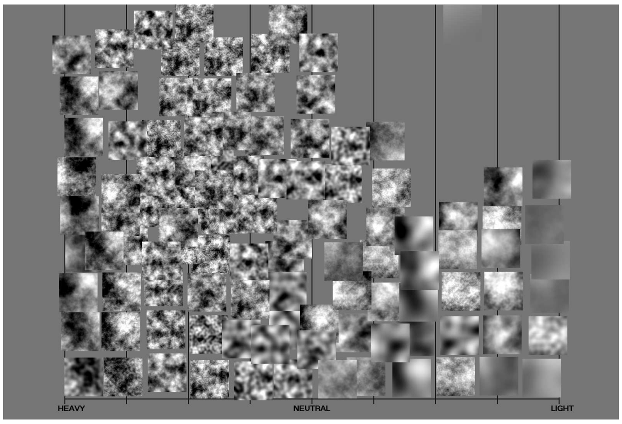

Figure 1b. *Experimental result for the Grayscale Texture samples, ordered horizontally along the Heavy – Light scale. Data from a single observer.*

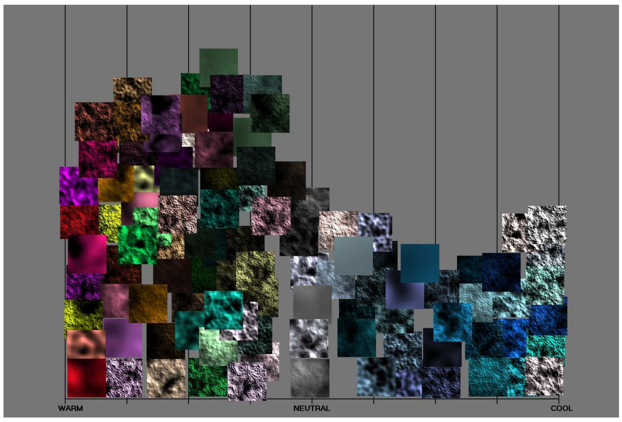

Figure 1c. *Experimental result for the Color Texture samples, ordered horizontally along the Warm – Cool scale. Data from a single observer.*

Intra-observer agreement

For each observer, sample type and emotion scale, we determined the rank correlation between the first and second measurement (Table 1). This correlation is a measure for the intra-observer agreement, or in other words, the repeatability. For 105 samples, the critical value of the correlation coefficient is about 0.195 at the 95% confidence level. Table 1 shows that the correlation between the first and second measurement is highly significant, for all subjects and all conditions, except for subject 6 on the Heavy-Light scale for the Grayscale Texture samples. For the Uniform Color samples the correlation averaged over subjects and emotion scales is 0.73, which is higher than the corresponding values for the Grayscale Texture samples (0.66) and the Color Texture samples (0.65). Apparently, subjects reproduce their color emotional responses on uniform samples better than on textured samples. Averaged over the three sample types, the highest intra-observer agreement is found for the Warm-Cool scale (r=0.74), followed by Heavy-Light (0.70), Masculine-Feminine (0.69) and Hard-Soft (0.60). Considering that the second measurement was made about one week after the first measurement, these intra-observer values seem satisfactory. Unfortunately, it is impossible to compare this result with other studies.

Table 1: Intra-observer agreement. Shown are the correlation coefficients between rank orders of the first and second measurement. WC=Warm-Cool, MF=Masculine-Feminine, HS=Hard-Soft, HL=Heavy-Light, avg=average.

Sample Type	Emo-scale	Subject										avg
		1	2	3	4	5	6	7	8	9	10	
Uniform Color	WC	0.73	0.83	0.78	0.74	0.81	0.82	0.76	0.78	0.79	0.79	**0.78**
	MF	0.82	0.72	0.77	0.79	0.74	0.92	0.76	0.79	0.40	0.65	**0.74**
	HS	0.86	0.67	0.42	0.32	0.66	0.66	0.33	0.77	0.86	0.56	**0.61**
	HL	0.83	0.87	0.87	0.90	0.85	0.89	0.54	0.75	0.93	0.60	**0.80**
	avg	**0.81**	**0.77**	**0.71**	**0.69**	**0.77**	**0.82**	**0.60**	**0.78**	**0.74**	**0.65**	**0.73**
Grayscale Texture	WC	-	-	-	-	-	-	-	-	-	-	-
	MF	0.87	0.79	0.76	0.57	0.61	0.67	0.67	0.68	0.85	0.38	**0.68**
	HS	0.73	0.82	0.71	0.78	0.72	0.76	0.80	0.68	0.87	0.58	**0.74**
	HL	0.66	0.54	0.69	0.62	0.76	0.10	0.68	0.42	0.50	0.47	**0.54**
	avg	**0.75**	**0.71**	**0.72**	**0.66**	**0.70**	**0.51**	**0.72**	**0.59**	**0.74**	**0.47**	**0.66**
Color Texture	WC	0.82	0.67	0.85	0.74	0.88	0.66	0.87	0.50	0.76	0.66	**0.74**
	MF	0.75	0.58	0.73	0.76	0.38	0.63	0.29	0.43	0.78	0.61	**0.60**
	HS	0.54	0.66	0.78	0.75	0.59	0.67	0.29	0.59	0.87	0.68	**0.64**
	HL	0.76	0.74	0.67	0.84	0.76	0.70	0.52	0.37	0.68	0.35	**0.64**
	avg	**0.72**	**0.66**	**0.76**	**0.77**	**0.65**	**0.67**	**0.50**	**0.48**	**0.77**	**0.57**	**0.65**

Inter-observer agreement

How well do observers agree with each other? In principle it is possible to calculate the correlation between rank orders for a single observer with each of the other observers, but this amounts to a lot of data. Instead we prefer to calculate the correlation between rank orders produced by each observer (averaged over the first and second measurement) and the average of all other observers in the group. This data is shown in Table 2.

Table 2: Inter-observer agreement. Shown are the correlation coefficients between rank orders of a single observer with the average of the other observers. WC=Warm-Cool, MF=Masculine-Feminine, HS=Hard-Soft, HL=Heavy-Light, avg=average. Avg=average.

Sample Type	Emo-scale	Subject										avg
		1	2	3	4	5	6	7	8	9	10	
Uniform Color	WC	0.41	0.83	0.90	0.20	0.82	0.27	0.59	0.51	0.47	0.73	0.57
	MF	0.91	0.81	0.79	0.76	0.86	0.85	0.74	0.83	0.73	0.54	0.78
	HS	0.77	-0.04	0.56	0.25	0.81	0.86	0.77	0.49	0.02	0.06	0.46
	HL	0.91	0.95	0.97	0.95	0.84	0.68	0.90	0.88	0.96	0.87	0.89
	avg	**0.75**	**0.64**	**0.81**	**0.54**	**0.83**	**0.67**	**0.75**	**0.68**	**0.55**	**0.55**	**0.68**
Grayscale Texture	WC	-	-	-	-	-	-	-	-	-	-	-
	MF	0.94	0.89	0.89	0.73	0.81	0.75	0.74	0.87	0.93	0.75	0.83
	HS	0.88	0.94	0.84	0.81	0.88	0.85	0.83	0.44	0.92	0.79	0.82
	HL	0.85	0.77	0.79	0.55	0.87	0.63	0.79	0.30	0.44	0.70	0.67
	avg	**0.89**	**0.87**	**0.84**	**0.70**	**0.85**	**0.74**	**0.79**	**0.54**	**0.77**	**0.75**	**0.77**
Color Texture	WC	0.63	0.87	0.79	0.11	0.81	0.63	0.82	0.78	0.63	0.60	0.67
	MF	0.79	0.80	0.78	0.45	0.71	0.73	0.59	0.60	0.33	0.22	0.60
	HS	0.53	0.88	0.76	0.79	0.54	0.83	0.67	0.64	0.78	0.69	0.71
	HL	0.77	0.83	0.28	0.80	0.79	0.56	0.84	0.06	0.78	0.70	0.64
	avg	**0.68**	**0.85**	**0.65**	**0.54**	**0.71**	**0.69**	**0.73**	**0.52**	**0.63**	**0.55**	**0.65**

Taking a look at the data in Table 2, we note that the average inter-observer correlation is 0.68 for the Uniform Color samples, 0.77 for the Grayscale Texture samples and 0.65 for the Color Texture samples, respectively. Apparently observers agree best on the Grayscale Texture samples. One salient result on the Uniform Color samples is that subjects 2, 4, 9 and 10 have low correlations with the group average on the Hard-Soft scale. This turned out to be partly attributable to the positioning of the dark samples along the scale. Further analysis showed that the standard deviation in the subject responses shows a minimum at L*+C*=100 and a more than two-fold increase at lower and higher values. Obviously, dark colors and saturated colors lead to lower agreement among subjects. This was found to apply to both the Warm-Cool and Hard-Soft scale.

It is impossible to state that the four observers indicated above did not do the experiment right, or that they should be regarded as outliers. Their correlation coefficients calculated between the first and second measurement (0.67, 0.32, 0.86 and 0.56, respectively) indicate that three of the four observers were able to replicate their results fairly well.

Before discussing the results of adding texture to the color samples we first present the results of regression analysis. This provides color emotion formulae with which we can more easily explain the effects of texture.

Quantitative analysis: color and texture emotion formulae

The goal of this section is to derive quantitative formulae that describe the color and texture emotions as a function of the samples' color and texture parameters. As a first step, we performed one-way ANOVA's to find out which of the parameters are significantly connected to the emotion scales. Using both the results of the one-way ANOVAs and formulas derived in previous studies [1] as a starting point, we derived analytical functions that gave the highest amounts of variance explained on the color emotion scales. The resulting functions are shown in Table 3. These functions predict the activity on the emotion scales, based on the color parameters L*,C*,h and/or the texture parameters octaves, frequency, persistence and lacunarity. The models were derived on the group averaged scale values. A negative scale value indicates a response towards the left word of the opposite word-pair (e.g. warm on the warm-cool scale), a positive value indicates a response towards the right word (e.g. cool on the warm-cool scale). A value of zero indicates neutral, i.e. exactly in the middle of the scale.

Table 3: Color and texture emotion formulae and percentages of explained variance. The adjusted R^2 measure already accounts for the number of free parameters in the formulae. The functions predict the activity on the emotion scales based on the CIELAB color parameters L*,C*,h and/or the texture parameters (oct=nr of octaves, freq=frequency, pers=persistence, lac=lacunarity). WC=warm-cool, MF=masculine-feminine, HS=hard-soft, HL=heavy-light. For the Color Texture samples, the table reports three values for the adjusted R^2. The first value relates to the function derived on the Color Texture samples, the second value relates to the function derived on the Uniform Color samples, and the third value relates to the functions derived for the Grayscale Texture samples, respectively.

Sample Type	Function for absolute scale values
UC	$WC_{UC} = -0.594 + 0.0166\,L - 0.209\,C^{0.6}\cos(h-45)$
	$MF_{UC} = -2.467 + 0.0346\,L + 0.802\,C^{0.3} - 0.0179\,h - 0.0000213\,h^2 + 0.000000228\,h^3$
	$HS_{UC} = -10.258 + 7.353\,L^{0.1} + 0.0527\,C - 0.00192\,C^2 + 0.0000108\,C^3 + 0.415\cos(h-30)$
	$HL_{UC} = -4.406 + 0.300\,L^{0.7} - 0.264\cos(h-130)$
GT	$MF_{GT} = 101.362 + 9.269\,L^{0.1} - 30.063\,oct^{0.05} - 6.059\,freq^{0.3} - 53.375\,pers^{0.1} - 25.151\,lac^{0.1}$
	$HS_{GT} = 116.121 + 6.103\,L^{0.1} - 32.295\,oct^{0.05} - 13.131\,freq^{0.1} - 48.807\,pers^{0.1} - 29.333\,lac^{0.1}$
	$HL_{GT} = 42.665 + 0.0643\,L - 12.459\,oct^{0.05} - 11.348\,freq^{0.1} - 5.844\,pers^{0.5} - 17.228\,lac^{0.05}$
CT	$WC_{CT} = -0.801 + 0.0151\,L - 0.2\,C^{0.65}\cos(h-40) + 0.0561\,oct$
	$MF_{CT} = 0.842\,L^{0.25} + 0.0215\,C - 0.0166\,h + 0.00000014\,h^3 - 0.566\,oct^{0.5} - 0.704\,freq^{0.5}$
	$HS_{CT} = 586.331 - 178.78\,oct^{0.01} - 84.202\,freq^{0.01} - 106.834\,pers^{0.02} - 213.891\,lac^{0.01}$
	$HL_{CT} = 0.334\,L^{0.6} + 0.0198\,C^{0.8} - 2.571\,oct^{0.1} - 1.414\,freq^{0.1} + 0.0153\,lac^3$

Sample Type	Function	Adjusted R^2	Avg R^2
Uniform Color (UC)	WC_{UC}	90.1	88.2
	MF_{UC}	83.2	
	HS_{UC}	81.5	
	HL_{UC}	98.1	
Grayscale Texture (GT)	MF_{GT}	82.5	82.3
	HS_{GT}	84.0	
	HL_{GT}	80.4	
Color Texture (CT)	WC_{CT}	84.4 / 81.5 / -	79.7
	MF_{CT}	75.8 / 39.6 / 0.0	
	HS_{CT}	72.5 / 0.0 / 48.8	
	HL_{CT}	86.2 / 58.7 / 0.0	

The Table shows that for the Uniform Color samples, the functions based on the CIELAB parameters L*, C* and h give rise to high values of adjusted R^2, with an average of 88.2. For the Grayscale Texture samples, the average is just 82.3. For the Color Texture samples, three values for the adjusted R^2 are reported. The first value relates to the function derived on the Color Texture samples, the second value relates to the function derived on the Uniform Color samples, and the third value relates to the functions derived for the Grayscale Texture samples, respectively. They show the relative importance of the color and texture parameters. On average, the adjusted R^2 for the Color Texture functions is 79.7.

All in all, the color and texture emotion functions provide a reasonably accurate description of the average observer responses. Uniform Color samples are best described, followed by Grayscale Texture and Color Texture. We now return to our main research question: what is the effect of texture on the color emotion scales?

The effects of texture on color emotion

We have already noted that the intra-observer agreement for the Uniform Color samples is higher than for the textured samples. At the same time, the inter-observer agreement is better for the grayscale samples than for the uniform samples and the color texture samples. The latter may be due to the fact that greyscale textures have no color, and so the observers have to deal with less dimensions.

The analytical functions presented in Table 3 reflect the dependencies on the samples' color and texture parameters. The color parameters L*, C* and h play an important role in all UC-functions, with the exception that C* does not appear in the function for the Heavy-Light scale. When looking at the functions for the Color Texture samples, we observe that 1) all color parameters L*,C* and h are present in the Warm-Cool and Masculine-Feminine scales, 2) only L* and C* appear in the Heavy-Light scale and 3) no color parameters appear in the function for the Hard-Soft scale. So, when texture is added to the uniform color samples, only the Hard-Soft scale loses its dependency on color parameters. In other words, Hard-Soft is fully dominated by texture. Warm-Cool, Masculine-Feminine and Heavy-Light are dominated by color parameters (in order of descending dominance), but need the texture parameters to explain for another 2.9%, 36.2% and 27.5% of the variance in the data, respectively.

Comparison with other studies

We can evaluate the performance of color emotion functions derived by others on our own experimental data, but only for the Uniform Color samples. Functions for grayscale or color texture samples have not been published previously. For the scales Warm-Cool, Hard-Soft and Heavy-Light we determined the adjusted R^2 for models derived by Sato et al. [14] and Ou et al. [1], see Table 4. The results show that our experimental data for the Heavy-Light scale (which heavily depend on lightness L*) is very well described by all three models. For the Warm-Cool scale, the model by Ou et al. [1] is good (R^2=0.70), but the model by Xin & Cheng [15] completely fails. For the Hard-Soft scale both models from Ou et al. and Xin & Cheng fail. An explanation for this may be the different methodologies used for obtaining the observer scores. In our experiments the subjects put the samples in relative order along the scale, whereas the other investigators only record the preference for one of the scale directions (for instance warm or cool). In the latter case, a final scale value is obtained by averaging over the scores of the observers, and therefore many observers are necessary.

Table 4: Performance (adjusted R^2) of color emotion models by different investigators on our experimental data for the uniform color samples.

Emotion Scale	Present study	Out et al. (2004)	Xin & Cheng (2000)
Warm - Cool	0.90	0.70	0.14
Hard - Soft	0.82	0.16+	0.36
Heavy - Light	0.98	0.96	0.96

+ excluding the 5 five achromatic samples. When including these samples (having C*=0), there is no correlation between our data and the model prediction by Ou et al. [1].

Discussion

We have demonstrated a systematic approach to the study of color emotions and the effect thereupon of adding texture to the color samples. A limited number of scales (four) was used because we were mainly interested in the specific effect of adding texture, and not so much in factor analysis that reveals how different scales may combine into new descriptors. Nevertheless, we have gathered a valuable set of experimental data using an improved methodology in which subjects ordered the samples along the scale while maintaining a view on all samples. Another methodological improvement in comparison to other studies is that our subjects repeated the experimental trials after one week which provided us with an estimate of the intra-observer agreement. We derived analytical functions that predict the group-averaged scale responses, with a precision exceeding that reported in other studies. The adjusted R^2 measure is the preferred measure to report, since that one corrects the R^2 for the number of free parameters in the functions.

Our subjects were from seven different nationalities. Testing on cross-cultural effects, as done in other studies, was not performed since that would require more subjects. Neither did we test on gender differences. Again, our focus was on the effect of adding texture, not on other issues. In the experimental design we adopted the minimum number of observers (10) as discussed in [16]). As long as the desired scale precision is unknown it is impossible to make precise estimates on the required number of observers. All that can be said is that the use of more subjects leads to lower standard deviations in the estimates. Scale accuracy increase with about the square root of the number of observers. Other studies have used more subjects (e.g. [1] used 31 observers, [2] used 70 observers, [3] used 50-70 observers per cultural group) but we preferred to perform a repetition of the full experiment, which we regard equally important. An interesting question is what the subjects' long term repeatability on the color and texture emotion scales is. That kind of information would greatly help to assess the validity and applicability of the color and texture formulae derived here.

References

[1] L.-C. Ou, M.R. Luo, A. Woodcock, A. Wright. A study of colour emotion and colour preference. Part I: Colour emotions for single colours. Color Res Appl, 29(3), 232-240 (2004).
[2] X.P. Gao, J.H. Xin. Investigation of human's emotional responses on colors. Color Res Appl; 31:411–417 (2006).
[3] X.P. Gao, J.H. Xin, T. Sato, A. Hansuebsai, M. Scalzo, K. Kajiwara, S-S. Guan, J. Valldeperas, M.J. Lis, M. Billger. Analysis of cross-cultural color emotion. Color Res Appl, 32(3), 223-229 (2007).
[4] C.E. Osgood, F.M. Adams. A cross-cultural study of the affective meanings of color. J Cross-Cultural Psychol;4:135–156 (1973).
[5] R. D'Andrade, M. Egan. The Colors of Emotion. Am Ethnol, 1(1), 49-63 (1974).
[6] T. Clarke, A. Costall. The emotional connotations of color: A qualitative investigation, Color Res Appl, 33(5), 406-410 (2008).
[7] M.A. Tinker. Effect of stimulus-texture upon apparent warmth and affective value of colors, Am J Psychol, 51(3), 532-535 (1938).
[8] S-J. Kim, E.Y. Kim, K.J. Jeong, J-I. Kim. Emotion-Based Textile Indexing Using Colors, Texture and Patterns. Proc. ISVC 2006, LNCS 4292, 9 – 18 (2006).
[9] L. Erhart, H. Irtel. Der Einfluss von Oberflächeneigenschaften auf die emotionale Wirkung von Farbe. FarbInfo, Dresden (2006).
[10] D.R. Simmons, C.L. Russell. Visual texture affects the perceived unpleasantness of colours. Perception 37 ECVP Abstract Supplement, p. 146 (2008).
[11] K. Perlin. An image synthesizer. Comp Graph, 19(3), 287-296 (1985).
[12] J.E. Bailey, M. Neitz, , D. Tait, J. Neitz. Evaluation of an updated HRR color vision test. Visual Neurosci. 21, 431-436, (2004).
[13] M. Stokes, M. Anderson, S. Chandrasekar, R. Motta. A standard default color space for the Internet–sRGB, November 1996, http://www.w3.org/Graphics/Color/sRGB.html.
[14] T. Sato, K. Kajiwara, H. Hoshino, T. Nakamura. Quantitative evaluation and categorising of human emotion induced by colour. Adv Colour Sci Technol;3:53–59 (2000).
[15] J.H. Xin, K. Cheng. Quantitative evaluation of colour emotion. JSPS Fund Meeting 2000;71– 86 (2000).
[16] P. Engeldrum. Psychometric Scaling: A Toolkit for Imaging Systems Development, Imcotek Press, Winchester, MA (2001).

Author Biography

Marcel Lucassen received an M.S. degree in Technical Physics from Twente University (The Netherlands) in 1988 and a Ph.D. in Biophysics (color constancy) from Utrecht University in 1993. In the period 1993-2007 he worked with Akzo Nobel Coatings and TNO Human Factors. He is now a freelance color scientist at Lucassen Colour Research and holds a part-time position at the University of Amsterdam. His interests lie in basic and applied vision research, and color vision in particular. He is an associate editor for Color Research and Application.