Colourfulness judgments of natural scenes

S.N. Yendrikhovskij *, H. de Ridder, E.A. Fedorovskaya ¹, F.J.J. Blommaert

Institute for Perception Research, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Received 25 October 1996; accepted 10 March 1997

Abstract

The purpose of this study was to examine the relationship between colourfulness judgments of images of natural scenes and statistical parameters of the chroma distribution over the images in the CIE 1976 L*u*v* (CIELUV) colour space. A comparative analysis of within-scenes and between-scenes colourfulness assessments was performed. Images were created by varying chroma in the CIELUV colour space while lightness and hue-angle were kept constant. Experiment 1 shows the results of a multidimensional analysis of the differences-scaling in perceived colourfulness. Experiment 2 includes a unidimensional investigation of the direct-scaling of the magnitude of perceived colourfulness. Both experiments confirm that colourfulness judgments highly correlate with a linear combination of the mean of the distribution of chroma values and its standard deviation. That was also verified in Experiment 3 with "Mondrian"-like stimuli (spatially scrambled versions of the images of natural scenes). Colourfulness judgments of natural scenes, but not "Mondrian"-like scenes, brought out systematic differences between observers. In order to interpret and quantify these differences, a model of individual strategies in colourfulness judgments is developed and discussed. © 1997 Elsevier Science B.V.

PsycINFO classification: 2323, 2260

Keywords: Color perception; Stimulus complexity; Multidimensional Scaling; Individual difference

---

* Corresponding author. Tel.: +31 40 2773877; fax: +31 40 2431930; e-mail: yendrik@ipo.tue.nl.

¹ Now at the Human Factors Laboratory, Eastman Kodak Co., Rochester, NY, USA.
1. Introduction

The concept of "colourfulness" has been introduced by Hunt (1977a) to specify the attribute of colour appearance accounting for the strength of the subjective chromatic response by which the hue is recognized. Colourfulness has been identified as the attribute of a visual sensation to which the perceived colour of an area appears to be more or less chromatic (CIE, 1970). For a given chromaticity, as the luminance increases, colourfulness generally increases (Hunt, 1977a). To obtain suitable measures that correlate with the magnitude of perceived colourfulness, Hunt has proposed the use of psychometric formulae of the CIE 1976 L*u*v* colour space (CIELUV). For a given level of adapting luminance, the colourfulness has been argued to be proportional to the CIELUV chroma (Hunt, 1977b). These measurements were obtained from experiments with simple patches on uniform backgrounds and have not yet been fully verified for real-life scenes with complex objects and backgrounds. For example, hardly anything is known about the correlation between perceived colourfulness and psychometric parameters of displayed images of natural scenes.

Recently, some progress has been made in modelling the relation between colourfulness judgments of displayed images of real-life scenes and statistical parameters of CIELUV chroma distributions over the images (Fedorovskaya et al., 1993, 1995, 1997). It has been proposed that subjective colourfulness depends on two factors. The first factor is the distance of the image colours from a neutral grey; this factor might be modelled as the mean of the distribution of the chroma values of all individual image colours. The second factor is the distance between individual image colours; this factor might be modelled as the spread of the average chroma expressed in, for instance, the standard deviation of the distribution of chroma values. Indeed, the experimental data have shown that direct-scaling of the perceived strength of colourfulness is linearly related to the weighted sum of the average chroma and the standard deviation of the chroma when those parameters were varied for original (unprocessed) images of four natural scenes. The direct-scaling of colourfulness, however, did not reveal any differences between the scenes, although the scenes had essentially different chroma distributions. The data suggest that observers made separate judgments for each scene. One can assume at least two reasons for this.

The first reason concerns the method used. There is abundant evidence that subjects' responses during the evaluation of an attribute varying within clearly distinguishable categories of stimuli depend on the experimental procedure, in particular the method of scaling (e.g., Parducci et al., 1976; Poulton, 1989). Results of direct-scaling of a given perceptual attribute usually characterize the magnitude variations only within categories. To describe variations between categories, we propose to use the method of difference-scaling.

The second reason refers to the specific aspect of natural scenes. The problem is that colour images of natural scenes include: (1) perceptual contents, i.e., complex colour patterns, and (2) cognitive contents, i.e., real-life objects. Both types of contents might allow for categorization. Apparently, categorization based on cognitive contents predominates in direct-scaling of an attribute of natural scenes causing observers to evaluate scenes separately. This means that while looking at a "landscape" scene,
for example, subjects will probably judge the colourfulness of "grass" rather than the
colourfulness of a "green" pattern (for an interesting discussion about distinguishing
attribute judgments from object judgments, see Lockhead, 1992). To separate the per-
ceptual aspect from the cognitive aspect of colourfulness judgments of natural scenes,
we propose to use a special type of images. These images, namely "micro-Mondrian",
include distribution of colour pixels in the CIELUV chromaticity plane similar to dis-
tribution of natural scenes but exclude any cognitive information.

The purpose of the present study is to further examine the relation between col-
ourfulness judgments of images of natural scenes and statistics of the chroma distri-
bution over the images in the CIELUV colour space. Particularly, we focus our at-
tention on a comparison of within-scenes and between-scenes colourfulness
judgments using two scaling methods. Experiment 1 presents the results of a multi-
dimensional analysis of the differences-scaling in perceived colourfulness. Experi-
ment 2 contains a unidimensional investigation of the direct-scaling of the magnitude
of perceived colourfulness. In addition, Experiment 3 with "Mondrian"-like stimuli
.spatially scrambled versions of the images of natural scenes) was carried out in or-
der to distinguish between the perceptual and cognitive aspects of colourfulness
judgments of natural scenes.

2. Experiment 1

2.1. Method

2.1.1. Subjects

Five subjects took part in the experiment. They had normal or corrected-to-nor-
mal vision. Their colour vision was checked with the H-R-R Pseudoisochromatic
Plates (Hardy et al., 1957). No significant colour deficiencies were observed.

2.1.2. Stimuli

Pictures of four natural scenes were chosen for the experiments: an abstract sculp-
ture with bushes ("Town hall"), an outdoor scene ("Terrace"), a portrait of a female
model ("Wanda"), and peaches displayed in front of a greengrocer’s shop ("Fruit").
Red, green and blue grey values for video signals were obtained by scanning the slide
and digitized with 8 bits per pixel on a grid of 512 x 512 pixel² using a Leaf System
Leafscan 35-SCSI. A region of interest of 450 x 450 pixel² was chosen for further
manipulations. The digitized image was described by its colour point distributions
in the CIELUV colour space through the sequential transformation of r, g, b grey
values to absolute R, G, B luminance values, then to the X, Y, Z tristimulus values,
and, eventually, to the L*, u*, v* colour coordinates. The transformation into the
CIELUV space was made by using standard formulae (Hunt, 1992). Each colour
point corresponded to 1 pixel. Reference white was D65 (Wyszecki and Stiles,
1982) with CIE chromaticity coordinates \((x_w, y_w) = (0.313, 0.329)\).

For every original image, a set of new images was computed by changing the chro-
ma value of each pixel while its lightness and hue-angle were kept constant. Two
ways of changing the chroma were used: (1) the same number of chroma units was added to or subtracted from the chroma value for each pixel (addition), and (2) the chroma value of each pixel was multiplied by a constant (multiplication). If, during the processing of the images, the calculated values for some pixels were outside the colour gamut of the monitor, the maximum possible value of chroma was used for those pixels (clipping). The complete experimental set included 20 images, five for each scene: two processed by addition, two processed by multiplication, and one unprocessed image (the original). Fig. 1 shows two images processed by multiplication, and one unprocessed image for the scene “Town hall” (Fig. 1, first column) and “Wanda” (Fig. 1, third column).

The images were displayed by an Image Sequence Processor ISP500 of Digital Video System DVS. A high-resolution 50 Hz non-interlaced BARCO CCID7351B colour monitor driven by a SUN-3/260 workstation was used to generate the stimuli. The monitor was colourimetrically characterized by using a Look-Up-Table calibration technique. Pixel luminance measured by Luminance meter LMT L1003 in a dark room ranged from 0.001 to 70 cd/m². The \( x, y \) chromaticity coordinates of red, green and blue phosphors, measured by Spectroradiometer SpectraScan PR-650, were \( (x_r, y_r) = (0.614, 0.349), \) \( (x_g, y_g) = (0.304, 0.594) \) and \( (x_b, y_b) = (0.149, 0.070) \), respectively.

Each stimulus on the monitor occupied an area of 21.5 × 21.5 cm² subtending of 8.24° × 8.24° of visual angle with a resolution of 55 pixels per degree of visual angle. The images were displayed for five seconds, while the interval between two image exposures lasted four seconds during which a 4.6 cd/m² homogeneous grey adaptation field appeared on the screen. All 20 stimuli were combined to form 400 stimulus pairs. Those pairs were presented in a random order once in two separate sessions, yielding four repetitions per pair, neglecting the order of presentation.

2.1.3. Procedure

The experiments were carried out in a dark room with a white dimly lit 2.5 cd/m² background behind the monitor. Subjects were seated approximately 140 cm from the screen. Other viewing conditions agreed with the CCIR Recommendation 500-2 (CCIR, 1992). After a five-minutes period of adaptation to room illumination and screen luminance, the subjects read the instructions and made a test series of 10 pairs to get familiar with the set of images and an experimental procedure.

The instructions stated that the purpose of the experiment was to investigate colourfulness, defined as the presence and vividness of colours in the whole picture. The subjects' task was to judge the difference in colourfulness between two successively presented images using integer numbers from 0 (no difference in colourfulness) to 10 (maximum difference in colourfulness).

2.2. Results

Two 20 × 20 matrices of colourfulness dissimilarity judgments were obtained for each observer. These matrices were combined into one averaged lower-half matrix (without the diagonal) per subject and processed by the multidimensional scaling
Fig. 1. Pictures of some images used in the experiments. Middle row: the unprocessed image of the "Town hall" scene (first column), the "micro-Mondrian" image of the "Town hall" scene (second column), the unprocessed image of the "Wanda" scene (third column), and the "micro-Mondrian" image of the "Wanda" scene (fourth column). Upper row: the same images as in the middle row with the CIELUV chroma multiplied by 0.8. Lower row: the same images as in the middle row with the CIELUV chroma multiplied by 1.2.
Symmetrical INDSCAL (SINDSCAL) program (Schiffman et al., 1981; Green et al., 1989). Multidimensional scaling methods generally attempt to describe a given set of stimuli as relations on points in a multidimensional space. The dimensions of the space are assumed to represent attributes along which the stimuli are compared. The SINDSCAL program determines, through an iterative least squares procedure, data by points in a n-dimensional space, such that the distances between the points are linearly related to the corresponding difference judgments. The advantage of this program is that it offers an opportunity of representing a common stimulus space which applies to all subjects, as well as, individual stimulus spaces for each observer.

One-, two-, three- and four-dimensional configurations were analysed by using the correlation coefficient $R_i$ between the estimated values and the scalar products derived from the dissimilarity matrix of subject i. According to this goodness-of-fit measure, a two-dimensional configuration was found to be the simplest solution providing an adequate fit for all individuals ($R_{es} = 0.82$, $R_{eh} = 0.84$, $R_{hr} = 0.88$, $R_{rj} = 0.86$, $R_{sy} = 0.94$).

The common stimulus space of the two-dimensional SINDSCAL solution is shown in Fig. 2(a). Projection of the stimuli on Dimension 1 includes both between-scenes and within-scenes variations, while projection of the stimuli on Dimension 2 includes only the within-scenes variation. To determine these variations, a model with two psychometrical parameters was developed.

The first parameter is defined as the “absolute”, i.e., independent, chroma-related value, $C_{abs}$:

$$C_{abs} = AVC_{is} + SDC_{is},$$

where $AVC_{is}$ and $SDC_{is}$ represent the average chroma and its standard deviation, respectively, for image i belonging to scene s. This parameter describes a simple version of the model proposed by Fedorovskaya et al. (1997) in which $AVC_{is}$ and $SDC_{is}$ have equal impact on colourfulness judgments. The second parameter is defined as the “relative”, i.e., reference-dependent, chroma-related value, $C_{rel}$:

$$C_{rel} = (AVC_{is} - AVC_{os}) + (SDC_{is} - SDC_{os}),$$

where $AVC_{os}$ and $SDC_{os}$ represent the average chroma and its standard deviation for the original image o, belonging to scene s, that is, the “relative” chroma-related values were zero-adjusted separately for each scene with respect to the unprocessed image. Later, the arbitrary choice of an unprocessed image as the reference will be discussed.

The “relative” chroma-related values are plotted in Fig. 2(b) as a function of the “absolute” chroma-related values. The relationship between those parameters has a similar pattern to the stimulus space of two-dimensional SINDSCAL solution. It appears that Dimension 1 highly correlates with the “absolute” chroma-related value, $C_{abs}$ (squared multiple $R^2 = 0.967$, $F = 531.7$, $p < 0.001$), and Dimension 2 correlates with the “relative” chroma-related value, $C_{rel}$ ($R^2 = 0.796$, $F = 70.3$, $p < 0.001$). However, comparing Fig. 2(a) and (b) reveals some differences.

Firstly, the range of “Town hall” images and, especially, the “Terrace” images in Fig. 2(a) are bigger than the model prediction in Fig. 2(b). This might be explained...
Fig. 2. (a) Two-dimensional stimulus space found by SINDSCAL for five subjects. (b) The model prediction of colourfulness judgments on the basis of “absolute” chroma-related and “relative” chroma-related values for four scenes: “Town hall” (□), “Terrace” (△), “Wanda” (○) and “Fruit” (◇).

by a general tendency of subjects to use the same response range and to neglect the physical range of the stimuli (Poulton, 1989). Secondly, the “Town hall” images in Fig. 2(a) are shifted slightly when compared with the images of other scenes: this may indicate that the subjects used a reference image which is slightly more saturated than the original. The results of Fedorovskaya et al. (1997) suggest that the reference image for the “Town hall” scene might be the image considered as the most “natural” one. Finally, despite the predictable small differences between “Fruit” and “Wanda” scenes, the order of these scenes along Dimension 1 for the experimental data is re-
versed when compared with the "absolute" chroma-related values of the model in Fig. 2(b). This might be due to other effects on the colourfulness judgments, e.g., lightness, and colour diversity (Fedorovskaya et al., 1995). The "Wanda" images, for instance, have average CIELUV lightness value of 44, and three clearly distinguishable colour clusters occupying different areas in the CIELUV colour space, i.e., yellowish face, pink shirt, and blue background. The reason might be that, in general, the "Wanda" images appear to be slightly more colourful than the "Fruit" images whose average CIELUV lightness value equals 37, and whose colours are more homogeneously distributed in the yellowish–reddish part of the CIELUV chromaticity plane.

3. Experiment 2

3.1. Method

The same five subjects as in Experiment 1, rated colourfulness on a ten-point numerical scale ranging from one (lowest colourfulness) to ten (highest colourfulness). The whole set of 20 images used in Experiment 1 was randomly presented five times in two sessions, resulting in 10 repetitions per image. Before starting the experiment, the subjects had to judge a test series of 20 images. In all other respects, Experiment 2 was identical to Experiment 1.

3.2. Results

Ten colourfulness judgments per image were obtained for each subject. To standardize the response range of subjects, individual data were transformed into z-scores. This linear transformation was produced by subtracting means and then dividing by standard deviations (Hays, 1988). The resulting colourfulness z-scores averaged over subjects are presented in Fig. 3(a) as a function of the "absolute" chroma-related value. Per scene, a nearly linear relation can be seen between the "absolute" chroma-related values and the colourfulness judgments. This supports the usefulness of the definition of $C_{abs}$ for the within-scenes variation. However, the correlation between colourfulness judgments and "absolute" chroma-related values over all scenes is low ($R^2 = 0.432$, $F = 13.685$, $p = 0.002$). The data show that the subjects used separate scales for each scene (compare Fig. 2(a) and Fig. 3(a)). Fig. 3(b) presents the colourfulness judgments as the function of the "relative" chroma-related values. Apparently, the curves for the four scenes are closer than in Fig. 3(a) which suggests that evaluations in the direct-scaling experiment tended to be based on the "relative" chroma-related values more than on the "absolute" ones. The results of a linear regression analysis demonstrate that the correlation between averaged colourfulness judgments and "relative" chroma-related values is statistically significant ($R^2 = 0.736$, $F = 50.223$, $p < 0.001$). The reason for the correlation being not very high is due to differences among the observers: EE ($R^2 = 0.831$, $F = 88.2$, $p < 0.001$), EH ($R^2 = 0.603$, $F = 27.386$, $p < 0.001$), HR ($R^2 = 0.748$, $F = 53.541$, $p < 0.001$), RJ ($R^2 = 0.724$, $F = 47.195$, $p < 0.001$), and SY ($R^2 = 0.394$, $F = 11.708$, $p = 0.003$).
Fig. 3. Colourfulness z-scores averaged over all subjects as a function of (a) “absolute” chroma-related values and (b) “relative” chroma-related values for four natural scenes: “Town hall” (■), “Terrace” (△), “Wanda” (○) and “Fruit” (◇). The vertical bars in the top-left corners indicate twice the average standard error of the mean of the judgments.

4. Experiment 3

4.1. Method

The images used in Experiment 1 were transformed in two ways. Firstly, in order to limit the influence of lightness on colourfulness judgments, the pixels of all images were set to the same CIELUV lightness value of 42, i.e., the average of all lightness
values for the images. Secondly, in order to limit the influence of cognitive content on colourfulness judgments, every image was divided into 8100 pieces of 5 x 5 size and intermixed using a scrambling algorithm to create “micro-Mondrian” images. Thus, the images in Experiment 3 had a similar distribution of colour pixels in the CIELUV chromaticity plane to the images in Experiments 1 and 2 (the distributions were not completely identical due to clipping), and were lacking of lightness variation and cognitive contents. Fig. 1 shows three examples of “micro-Mondrian” images for the scene “Town hall” (Fig. 1, second column) and “Wanda” (Fig. 1, fourth column).

Five subjects, only two of whom participated in Experiments 1 and 2, had to judge colourfulness on a eleven-point numerical scale ranging from zero (lowest colourfulness) to ten (highest colourfulness). The whole set of 20 images was randomly presented twice in two sessions, resulting in four repetitions per image. In all other aspects, Experiment 3 was identical to Experiment 2.

4.2. Results

Four colourfulness judgments per image were obtained for each subject. The averaged colourfulness z-scores are presented in Fig. 4 as functions of the “absolute” chroma-related values. A clear linear relationship between the psychometrical and perceptual attributes can be observed. The results of a linear regression analysis demonstrate that the correlation between colourfulness judgments and “absolute” chroma-related values is very high for averaged data ($R^2 = 0.987$, $F = 1357$, $p < 0.001$), and per subject: LY ($R^2 = 0.945$, $F = 312$, $p < 0.001$), NB ($R^2 = 0.956$, $F = 394$,

![Graph showing colourfulness z-scores averaged over all subjects as a function of “absolute” chroma-related values for four “micro-Mondrian” scenes: “Town hall” (□), “Terrace” (△), “Wanda” (○) and “Fruit” (◇).](image-url)
5. A model of individual strategies in colourfulness judgments

In order to interpret and quantify the individual differences in Experiment 1 and 2 the following model was developed. The basic assumption of the model is that, in an experiment with several scenes, the colourfulness judgments of any particular image are based on a combination of two strategies: (1) using one unified scale and making unified judgments according to "absolute" chroma-related values of the image, and (2) using different scales for each scene and making separate judgments according to the "relative" chroma-related values. The colourfulness rating, \( R_{is} \), of image \( i \) belonging to scene \( s \) represents a weighted average of these two strategies:

\[
R_{is} = k[wC_{is(abs)} + (1-w)C_{is(rel)}] + c,
\]

where \( C_{is(abs)} \) is the "absolute" chroma-related value of image \( i \) belonging to scene \( s \), \( C_{is(rel)} \) is the "relative" chroma-related value of the same image, and \( k \) and \( c \) are coefficients of the linear mapping between the stimulus and response scales. The weighting parameter, \( w \), describes a compromise between the two strategies: when \( w = 1 \), the judgments are based on the "absolute" chroma-related values of the image only, and when \( w = 0 \), the judgments are based on the "relative" chroma-related values of the image only. The "relative" chroma-related value, \( C_{is(rel)} \), is defined as follows:

\[
C_{is(rel)} = C_{is(abs)} - C_{s(ref)}.
\]

Hence,

\[
R_{is} = k[wC_{is(abs)} + (1-w)(C_{is(abs)} - C_{s(ref)})] + c,
\]

\[
R_{is} = k[C_{is(abs)} + (w-1)C_{s(ref)}] + c,
\]

where \( C_{s(ref)} \) is the "absolute" chroma-related value of the reference for scene \( s \). The complete colourfulness rating is derived from Eqs. (1), (2) and (6) as follows:

\[
R_{is} = k[(AVC_{is} + SDC_{is}) + (w-1)(AVC_{os} + SDC_{os})] + c.
\]

The model was tested by linear regression analysis for the results of Experiment 2. The value of \( R^2 \) for colourfulness judgments averaged over all subjects is 0.903 (\( F = 79.6, p < 0.001 \)) demonstrating a high correlation between the data and the model. Fig. 5 represents the averaged colourfulness z-scores for Experiment 2 as a function of the combined "absolute" and "relative" chroma-related values obtained by regression analysis. "Combined" chroma-related values, \( C_{combined} \), are defined as:

\[
C_{combined} = 0.19*C_{is(abs)} + 0.81*C_{s(ref)}.
\]

It can be seen from Fig. 5 that curves for the four scenes are closer together than in Fig. 3(a) or Fig. 3(b).

In this model it is assumed for simplicity that \( C_{s(ref)} \) is the "absolute" chroma-related value of the original image (see Eq. (2)). However, it is obvious that this does
not necessarily have to be the case. The reference image is the general expression for the ideal or prototype image that represents a certain category. The ideal image may vary per individual and may be, for example, “the most natural” one. This might explain the shift found in Experiment 1 along Dimension 2 for the “Town hall” scene, where the most natural image was considered to be more colourful than the original (Fedorovskaya et al., 1997). Note, also, that instead of the scenes one might mean any subset of the complete collection of images which are categorized by observers into a coherent whole: it also can consist of a number of exemplars from the same category which can be easily distinguished by some criterion, e.g., size, periodic structure, blur, noise, and others.

The weighting parameter that describes the compromise between the two strategies was calculated per subject for the difference-scaling experiment. It has been noted that the SINDSCAL program determines individual stimulus spaces for each observer. The individual spaces are related to the common space by weighting the dimensions of the common space. The weighting coefficients define sensitivity to variation along dimensions and can be used to calculate the weighting parameter, $w$, in the model:

$$w = 2 \arctan\left(\frac{d_{2j}}{d_{1j}}\right)/\pi,$$

where $d_{1j}$ and $d_{2j}$ are the weighting coefficients of Dimensions 1 and 2 for subject $j$ provided by the SINDSCAL program. The weighting parameters, $w$, for the subjects participating in Experiment 1 are shown in Fig. 6(a). The weighting parameters, $w$, for the subjects participating in Experiment 2 were calculated with Eq. (7) using linear regression analysis and they are presented in Fig. 6(b). For all subjects who took part in Experiment 3, the weight for the “absolute” judgments was close to 1 and for the “relative” judgments was not significant at the 0.05 level, thus $w$ was considered to be 1.
Fig. 6(a) demonstrates that, in the case of the difference-scaling procedure, observers largely based their judgments on "absolute" chroma-related values (average \( w = 0.73 \)). Some observers, RJ \((w = 1.00)\) and SY \((w = 0.98)\), based their judgments only on "absolute" chroma-related values, while others, EE \((w = 0.72)\), EH \((w = 0.59)\), and HR \((w = 0.35)\), used both "absolute" and "relative" chroma-related values. In the case of the direct-scaling procedure (Fig. 6(b)) with natural scenes, observers largely based their judgments on "relative" chroma-related values (average \( w = 0.19 \)). Some of the subjects based their judgments only on these values.
(w = 0.06 for EE; w = 0.05 for HR), while, others also paid attention to the “absolute” chroma-related values (w = 0.29 for EH; w = 0.20 for RJ; w = 0.40 for SY). No significant subjective differences were found in Experiment 3 where all subjects judged colourfulness in relation to “absolute” chroma-related values.

6. General discussion and conclusions

The present paper confirms that colourfulness judgments of images of natural scenes can be derived from a basic statistical analysis of the chroma distribution in the CIELUV colour space. Colourfulness highly correlates with a linear combination of the mean of chroma and the standard deviation of chroma. Thus, when people look at a complex image, two primary factors shape their subjective impression of how colourful an image appears. The first factor corresponds to the average chroma value of all pixels in the images. The second factor corresponds to the variation of this average chroma value. The higher the values of these parameters the more colourful the image is judged to be. This holds for both within-scenes and between-scenes estimations. The relation between colourfulness judgments and statistics of the chroma distribution was also verified with a new type of scene, namely “micro Mondrian”, which has been developed for Experiment 3 of this paper.

It is important to emphasize two characteristic properties of the “micro-Mondrian” scenes. Firstly, their distributions of the colour pixels in the CIELUV chromaticity plane are almost identical to those of the real-life images resulting in a more ‘complex’ composition than simple-patches or ordinary Mondrian-like images. The fact that for both the natural and “micro-Mondrian” scenes, the observer’s judgments can be described by essentially the same parameter, i.e., the sum of the mean and the standard deviation of chroma in the CIELUV colour space, validates the use of this parameter for predicting and calculating perceived colourfulness of complex patterns. Secondly, the “micro-Mondrian” scenes have no cognitive contents and, thus, have less ‘natural’ composition than real-life scenes. The absence of meaningful information limits the possibility of cognitive categorisation and focuses the subject’s attention on the perceptual categorisation of images. As the result of this limitation, and in contrast with Experiment 1 and 2, no significant subjective differences were found in Experiment 3 where all subjects followed one strategy to judge colourfulness in relation to “absolute” chroma-related values. In general, we suggest that experiments with “micro-Mondrian” images might link simple-patches experiments with natural-scenes experiments.

The results of this study bring up the problem of subjective strategy in scaling methods. We assume that in quantifying judgments observers are sometimes put in a dilemma: on the one hand, they are supposed to base their judgments on real stimulus magnitude; on the other hand, they should respond as accurately as possible. The second demand is represented by response-range equalizing bias: whatever the range of responses is that subjects are given, they tend to use most of the range to describe the stimuli (Poulton, 1989). As long as a set of stimuli is not very large, these demands are not contradictory. However, in an experiment with a large set of differ-
ent groups of stimuli, it is difficult to use a single unified scale. Some subjects might follow another strategy: they use a separate scale for each stimulus category. The larger the set of stimuli the observers have to deal with, the more obvious this strategy appears to all of them. This explains why no individual differences were found in a previous study of colourfulness (Fedorovskaya et al., 1993, 1997) where the whole set of stimuli consisting of 84 images was much larger than that in the present study. Thus, if an instruction in a scaling experiment does not provide information about whether observers should use one scale or separate scales for each group of stimuli, both strategies should be considered.

The problem of subjective strategy in colourfulness judgments was solved by making a comparison between the results from direct- and difference-scaling experiments. We assume that both scaling procedures must be based on the same perceptual foundation and might be based on different decision-making strategies. The data support this assumption and demonstrates that the difference-scaling procedure corresponds to the use of one unified scale, and the direct-scaling procedure corresponds to the use of separate scales. This conclusion agrees with results of Van Dijk et al. (1995) who showed that when assessing the perceived quality of images comprising one of two different coding artifacts, observers changed their decision to use separate scales during the direct-scaling experiment to one unified scale during the difference-scaling experiment. However, the change of strategy did not take place at the same rate for all observers, which again stresses the importance of individual data analysis in attribute judgments of natural scenes.

Acknowledgements

The authors wish to thank Prof. J.A.J. Roufs and Prof. Ch.M.M. de Weert for their helpful comments and valuable advice.

References