Image similarity for chromatic content

M. Scheller Lichtenauer[1,3], J. Preiss[2], P. Urban[2], P. Zolliker[1]

[1]EMPA
Abteilung
Medientechnik
Ueberlandstrasse 129,
8600 Duebendorf
Schweiz
http://empamedia.ethz.ch

[2]Technische Universität Darmstadt
Institut für Druckmaschinen
und Druckverfahren
Magdalenenstr. 2,
64289 Darmstadt
Deutschland
http://www.idd.tu-darmstadt.de

[3]Friedrich Schiller Universität
Fakultät für Mathematik
und Informatik
Ernst-Abbe-Platz 2,
07743 Jena
Deutschland
http://theinf2.informatik.uni-jena.de

{matthias.scheller, peter.zolliker}@empa.ch, {preiss, urban}@idd.tu-darmstadt.de

Abstract Image similarity describes the quality of an image relative to a reference. Image difference measures are calculated in order to quantify the perceived difference. The concept is also applicable to model *where* two colour images differ. Existing measures for image similarity focus on high spatial frequency distortions, best visible in the luminance channel. Lower frequency distortions resulting from gamut mapping, tone mapping or illuminance change are only incorporated as far as they influence luminance. As luminance reduction and chroma reduction are often correlated in real applications, we studied image similarity on artificial transformations, thus allowing us to reduce luminance or chrominance independently of each other.

1 Introduction

We recently presented an image difference measure incorporating lower frequency chromatic distortions [1, 2, 3]. Although the data basis we could test these measures on is quite big, including data from distortions like noise, blur, compression or gamut mapping, they lacked several properties. First, noise in that data is of high frequency. Artefacts of JPEG compression tend to be of high frequency nature, too, as that algorithm uses blocks of 8×8 pixels and block borders become visible at higher compression factors and dominate over chromatic changes. In gamut mapping studies, chromatic changes were often correlated with luminance changes. Furthermore, effects of memory colours could not be studied. Enough reasons to conduct a study with synthetic distortions allowing to compare perception of both, abstract images and pictures of natural scenes.

2 Previous work

While colour difference measures are mostly based on nearly identical, uniform, structureless patches on an achromatic background [4], image difference measures have to take the spatially varying character of the stimuli into account. Spatial characteristics of contrast sensitivity have been investigated with regular structures known as gratings or Gabor patterns. Human ability to discriminate luminance contrasts in gratings is maximal at about 3–6 cycles per degree of visual angle and diminishes at higher and lower frequencies [5, 6]. Sensitivity to isoluminant chrominance contrast (i.e. red-green or blue-yellow gratings) is higher than to luminance gratings below 0.3 cycles per degree, but lower at higher spatial frequencies [7]. The subject of fading chromatic contrast with small visual angle is known from research on legibility and visual search [8, 9]. In order to account for spatial contrast sensitivity, a filter with spatial characteristics appropriate to the pixel size and viewing distance can be applied to the images prior to calculate difference metrics [10].
One of the best accepted image difference measures, named structural similarity index measure (SSIM), was initially calculated on a single scale [11], but later extended to multiple scales [12]. Though not identical to contrast sensitivity filtering, using different scales makes the measure more robust to fine distortions that may not be visible at a larger distance. SSIM is, however, formulated and tested by its authors only on the luminance channel of images.

3 Experimental

We chose ten digitised abstract images painted by Paul Klee [13] with mostly highly chromatic content as well as ten natural images. We included three types of distortions, which were simultaneously applied with varying degree to the images:

- reduction in chroma, leaving lightness computationally untouched
- simulated reduction in exposure time of a photography
- simulated effects of flare on displays

To minimise unintended hue shifts, the first distortion was calculated in the hue linear colour space proposed by Lissner and Urban [14], while the latter two distortions were calculated in CIEXYZ. The code is listed in the Appendix. Ten experienced observers with normal colour vision passed a total of 2050 paired comparisons, each facing two distorted images with the same reference image on an EIZO CG220 LCD display (Presentation used Psychtoolbox [15]). They answered, which distortion resulted in a more accurate reproduction of the reference. If the two distorted images were perceived identical, observers were instructed to reject the decision; so they did in 52 of 2050 trials.
Reference images were randomly selected. For half of the comparisons, distortion parameters were as well randomly selected. For the other half of the comparisons, the algorithm of [16] was adapted to actively select the parameters out of 10000 randomly created parameter sets. Active sampling should focus trials on border cases.
The active sampling algorithm builds a tree with the random parameter sets as leaves. Each leave has a strictly positive probability to be sampled. The random choice of a parameter set is restricted to a subtree in each trial. For each subtree, a score is calculated based on the number of leaves contained in it, past choices within the subtree

and confidence estimations. Active sampling then uses said score to chose a subtree and balance between the goals of consistency and accuracy.

Apart of being used to further test image similarity measures, the recorded choices allow us to compare perception of abstract and natural images with regard to the distortions, but also the random and active sampling strategy.

4 Results

4.1 Image difference measure

We first assessed the performance of two image similarity measures. A basic, luminance-only version of SSIM [11] was compared to a colour image difference measure (CID) combining contrast sensitivity filtering, SSIM features and chromatic features. For the implementation, please refer to the Appendix.
To assess the performance, we used the fraction of correctly predicted choices, called hit rate. We separately calculated these fractions for all comparisons involving a particular reference image. Hit rate of CID was better than that of SSIM for each reference image. Visually, the difference is significant (Figure 1). The boxes in all plots span from the 25% percentile to the 75% percentile, the median is marked with a red line. The whiskers extend to the most extreme data not considered to be outliers and outliers are marked with crosses. Neither SSIM nor CID resulted in significantly different performance for abstract or natural images (Figure 2).

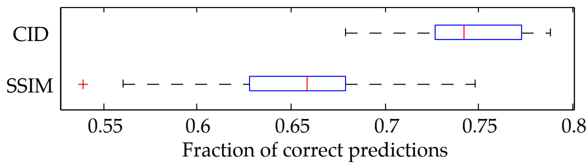

Figure 1: Performance of SSIM [11] compared to CID. All choices involving the same reference image contributed one single point to the hit rate distribution.

4.2 Machine learning

We validated the results with regard to perception of distortions on natural vs. abstract images with a completely different approach, namely machine learning methods. The machine learning aproach had no access to image information as had the similarity measures, but was trained on the parameters of the image distortion algorithm. By using two so different methods we can exclude that our conclusions will be due to the prediction method used. The parameters of the distortion algorithm for both distorted images were fed to a polynomial kernel (d=3) support vector machine [17, 18]. The five parameters of the

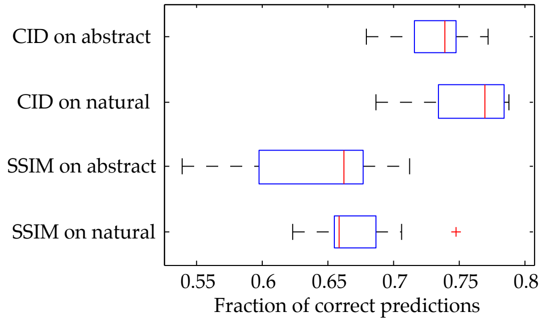

Figure 2: The prediction of choices on natural images was not significantly different than on abstract images.

distortion algorithm are documented by the source code in the appendix. Parameters did not include the image data, but a binary flag indicating whether the original image was abstract or natural was added as a meta-parameter [19]. The regularisation parameter of the support vector machine was chosen with leave-one-out cross validation.

As we did with the similarity measures, we calculated the fraction of correctly predicted choices on all comparisons involving a particular reference image. Results are shown in Figure 3. One should note that in the top panel of Figure 3, the choices to be predicted were sometimes included in the training set (e.g. learn random-predict random), sometimes not (e.g. learn random - predict active), while in the lowest two panels, choices to be predicted were never present in the training set. Performance on subsets of the training set is usually higher than on a completely independent test set, so the respective differences in median of distributions in the top panel of Figure 3 are primarily due to properties of learning methods.

We additionally tested, whether selecting subsets by choosing trials involving distortions from one particular reference image would result in a different hit rate distribution than when randomly choosing subsets. For this, we randomly split all choices from abstract and random sampling in two halves, trained a machine on one half and predicted choices on randomly chosen subsets of size $n = 50$ from the other half — this corresponds to the expected size of subsets when dividing by reference images (Figure 3, lowest panel).

5 Discussion and Conclusions

Our primary goal was to compare perception of distortions on abstract and natural images. We used artificial distortions, since chromatic distortions in gamut mapping often tend to be correlated in luminance and chroma. The artificial distortions did not modify grey balance, but one of three distortions reduces chroma while leaving luminance computationally untouched.

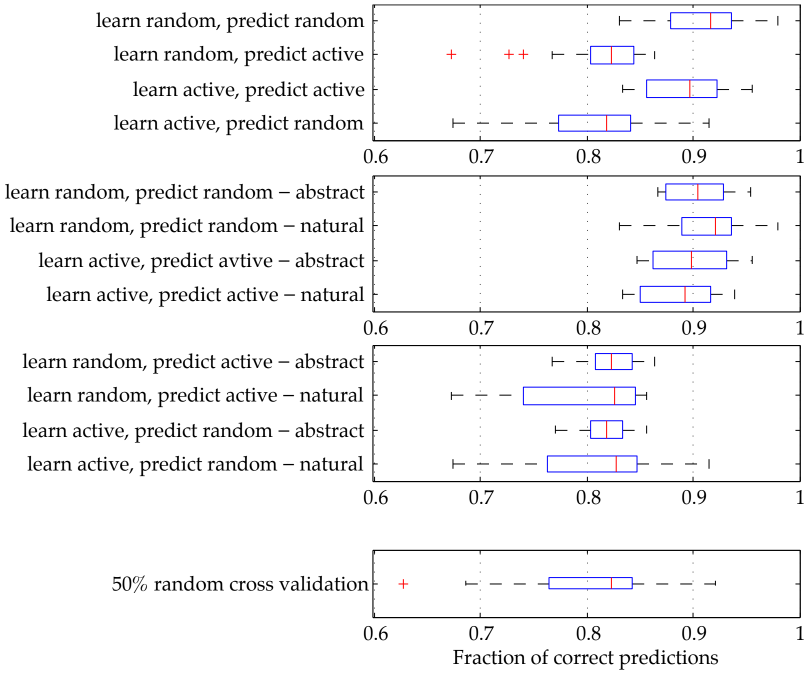

Figure 3: Predicting choices with polynomial kernel support vector machines (d=3) trained on different subsets of identical size. Note that in some cases, predictions are on a subset of the training set, in others not. One generally expects lower performance in the latter case.

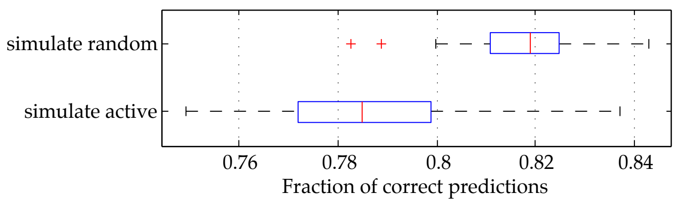

Figure 4: Simulating active vs. random sampling.

Predictions on distorted natural images do not significantly differ from predictions on abstract images for our sample of reference images. This main result is supported by both approaches, image difference measures as well as machine learning. This is remarkable, since the first approach had access to the images, but no information on parameters of the distortion algorithm, while the latter approach had exactly the opposite information to base the prediction on.

As in previous work on the subject of distortions in colour images [1, 2, 3], we again observe higher hit rates in predicting choices when using chromatic features than when using the achromatic channel only, i.e. our CID with contrast filtering is superior to the used basic version of SSIM with regard to that task. This agrees with results presented in [3] for newer and more sophisticated versions of SSIM.

The results with regard to the active and the random sampling method are not conclusive. We analysed sampling methods with predictions from the machine learning approach only. We observe the usual drop in performance between subsets of the training set relative to subsets of the test set. The spread of predictions on abstract and natural images is not the same, but the median is not significantly different. Last, but not least, we compared the active with the random sampling strategy with a simulation. Both strategies could chose 999 data points from the 1998 non-tied responses we had in the experiment. Both sampling strategies had then to predict the other choices. However, the active strategy frequently chose to query a data point more than once. In a real experiment, the answers might be different each time, but not in the simulation. Thus, the difference in hit rate that is visible in the results presented in Figure 4 could be attributed to this bias.

Appendix: Calculations

Calculation of a basic Structural Similarity Index Measure (SSIM)

Let x indicate a window around a pixel position in the original image X, and let y be a corresponding window in the distorted image Y. The SSIM as we use it reads as follows:

$$\mathrm{SSIM}(X,Y) = \overline{\mathrm{l}(x,y) \cdot \mathrm{c}(x,y) \cdot \mathrm{s}(x,y)} \quad , \tag{1}$$

with over-line indicating averaging over all windows and the following luminance feature functions per window:

$$\mathrm{l}(x,y) = \frac{(2\mu_x\mu_y + c_1)}{(\mu_x^2 + \mu_y^2 + c_1)} \quad , \tag{2a}$$

$$\mathrm{c}(x,y) = \frac{(2\sigma_x\sigma_y + c_2)}{(\sigma_x^2 + \sigma_y^2 + c_2)} \quad , \tag{2b}$$

$$\mathrm{s}(x,y) = \frac{(\sigma_{xy} + c_3)}{(\sigma_x\sigma_y + c_3)} \quad . \tag{2c}$$

The constants are set to the values in [11], namely $c_1 = (0.01L)^2$, $c_2 = (0.03L)^2$ and $c_3 = \frac{1}{2}c_2$. L is the numeric dynamic range of the pixel values e.g. 255 for 8-bit RGB. The symbols μ_x and σ_x denote empirical mean and standard deviation in the sliding window of 9×9 pixels, σ_{xy} is the covariance between corresponding windows.
More sophisticated versions of SSIM are documented in [11, 12]

Calculation of Colour Image Difference Measure (CID)

The source code of the calculations for the colour image difference measure (CID) can be found at: http://www.idd.tu-darmstadt.de/color/papers as 'Supplementary material'. Therefore, we only shortly sketch the main points here.

The images are pre-filtered to account for the contrast sensitivity of the human visual system as in [10]. Pixel resolution at the viewing distance was estimated to be 20 cycles per degree in average, corresponding to 70cm viewing distance to the display. An implementation of the filtering is provided as well in the source code.

Compared to SSIM, additional features modelling hue shifts [$\mathrm{h}(x,y)$] and chroma shifts [$\chi(x,y)$] are used to calculate the colour image difference measure:

$$\mathrm{CID}(X,Y) = 1 - \overline{\mathrm{l}(x,y)} \cdot \overline{\mathrm{c}(x,y)} \cdot \overline{\mathrm{s}(x,y)} \cdot \overline{\chi(x,y)} \cdot \overline{\mathrm{h}(x,y)} \quad . \tag{3}$$

SSIM describes similarity, hence the prediction is 1 if the two images compared are identical. CID describes a perceptual difference, hence the prediction is 0 if the two images compared are identical. As the individual feature functions are derived from SSIM, the '1-' in above equation accounts for predicting difference rather than similarity

The luminance features (l, c and s) of the colour image difference measure describe mean luminance, contrast and correlation of deviations respectively as do those above for SSIM, but they are calculated in a uniform colour space as documented in the appendix of [3].

The contribution of the individual pixels i in the sliding windows of 11×11 pixels are weighted with a circular-symmetric Gaussian weighting function ($\sigma = 1.5$ pixels, $\sum w_i = 1$) centred in the middle of the window around x respectively y. The luminance feature function $\mathrm{l}(x,y)$ per window then becomes:

$$\mathrm{l}(x,y) = \frac{1}{c_1 \cdot \sum w_i \cdot \Delta\mathrm{L}(x_i,y_i)^2 + 1} \quad . \tag{4}$$

Calculation of the contrast feature function $\mathrm{c}(x,y)$ and the structure feature function $\mathrm{s}(x,y)$ does not change except for the Gaussian weighting.

The two additional features for chroma and hue are set up with the following feature functions per window, with pixels in the respective window indexed by i:

$$\chi(x,y) = \frac{1}{c_4 \cdot \sum w_i \cdot \Delta\mathrm{C}(x_i,y_i)^2 + 1} \quad , \tag{5}$$

$$\mathrm{h}(x,y) = \frac{1}{c_5 \cdot \sum w_i \cdot \Delta\mathrm{H}(x_i,y_i)^2 + 1} \quad . \tag{6}$$

The hue and chroma differences $\Delta\mathrm{H}$ and $\Delta\mathrm{C}$ can be calculated from Cartesian chromaticity coordinates a and b as follows:

$$\Delta\mathrm{C}(x_i,y_i) = \sqrt{a_{x_i}^2 + b_{x_i}^2} - \sqrt{a_{y_i}^2 + b_{y_i}^2} \quad , \tag{7}$$

$$\Delta\mathrm{H}(x_i,y_i) = \sqrt{(a_{x_i} - a_{y_i})^2 + (b_{x_i} - b_{y_i})^2 - \Delta\mathrm{C}(x_i,y_i)^2} \quad . \tag{8}$$

For calculations in this paper, the parameters were set as follows:

c_1	c_2	c_3	c_4	c_5
0.002	0.1	0.1	0.002	0.008

As a last remark, the calculations have been performed in the hue linear colour space LAB2000HL [14]. The source code necessary for transformation of sRGB images from and to this space can be found at:
http://www.idd.tu-darmstadt.de/color/papers as 'Supplementary material' of [14].

The hue linear colour space has as well been used in calculation of the distortions documented as source code in the following. All source code uses MATLAB as a platform.

We provide images used in the experiments documented here to interested researchers. Please contact the primary author.

Calculation of distortions

```
function [ tim ] = mytransform( im, lfactor, afactor, bfactor,...
                   ang, flare)
% 2012/03/23 by Matthias Scheller Lichtenauer
% im is expected to be n x m x 3 in [0,1]
% lfactor, afactor, bfactor, flare in [0,1], ang in [0,pi]
% Based on SRGB2LAB2000HL() and LAB2000HL2LAB()
% 2011 by Philipp Urban and Ingmar Lissner

% Reshape the data if necessary
if (ndims(im)==3 && size(im,3)==3)
   d1=size(im,1);
   d2=size(im,2);
   im=reshape(im,d1*d2,3);
end

%transform to LAB2000 hue linear space
tim=SRGB2LAB2000HL(im);

% Rotate around L*-axis and compress so that
% circles in chromaticity planes are mapped to smaller ellipses
% (assumed invariant in lightness)
rotim2=afactor*(tim(:,2)*cos(ang)-tim(:,3)*sin(ang));
rotim3=bfactor*(tim(:,2)*sin(ang)+tim(:,3)*cos(ang));
% Undo rotation
tim(:,2)=(rotim2*cos(-ang)-rotim3*sin(-ang));
tim(:,3)=(rotim2*sin(-ang)+rotim3*cos(-ang));

%Simulate reduction in exposure in XYZ
tim=lfactor*LAB2XYZ(LAB2000HL2LAB(tim));

%Simulate achromatic flare in XYZ
fla=SRGB2XYZ(flare*[32 32 32]./256);
tim(:,1)=(1/(1+fla(1)))*(tim(:,1)+fla(1));
tim(:,2)=(1/(1+fla(2)))*(tim(:,2)+fla(2));
tim(:,3)=(1/(1+fla(3)))*(tim(:,3)+fla(3));

%transform to sRGB [0..255]
tim=uint8(XYZ2SRGB(tim)*255);
%reshape the data if necessary
if (ndims(im)==3 && size(im,3)==3)
   tim=reshape(tim,d1,d2,3);
end
end
```

Acknowledgements

This research was supported by the Swiss National Science Foundation and the German Research Foundation.

References

[1] M. Scheller Lichtenauer, P. Zolliker, I. Lissner, J. Preiss, and P. Urban. Learning image similarity measures from choice data. In *6th European Conference on Colour in Graphics, Imaging, and Vision*, pages 24–30. IS&T, Springfield, VA, 2012.

[2] J. Preiss, I. Lissner, P. Urban, M. Scheller Lichtenauer, and P. Zolliker. The impact of image-difference features on perceived image differences. In *6th European Conference on Colour in Graphics, Imaging, and Vision*, pages 43–48. IS&T, Springfield, VA, 2012.

[3] I. Lissner, J. Preiss, P. Urban, M. Scheller Lichtenauer, and P. Zolliker. Image-Difference Prediction: From Grayscale to Color. *IEEE Transactions on Image Processing*, accepted 2012.

[4] G. Wiszecki and W.S. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulae.* John Wiley & Sons, Inc., 1982.

[5] J.G. Robson. Spatial and temporal contrast-sensitivity functions of the visual system. *J. Opt. Soc. Am.*, 56(8):1141–1142, Aug 1966.

[6] F.W. Campbell and J.G. Robson. Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, 197:551–566, 1968.

[7] K.T. Mullen. The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *Journal of Physiology*, 359:381–400, 1985.

[8] Thomy Nilsson. Legibility of colored print. In *International encyclopedia of ergonomics*, volume I, chapter 293, pages 1440–1452. CRC Press, New York, 2006.

[9] Robert Carter and Rafael Huertas. Ultra-large color difference and small subtense. *Color Research & Application*, 35(1):4–17, 2010.

[10] X. Zhang, B.A. Wandell, et al. A spatial extension of CIELAB for digital color image reproduction. *SID international symposium digest of technical papers*, 27:731–734, 1996.

[11] Z. Wang, A. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4):600–611, 2004.

[12] Z. Wang and A Bovik. Mean square error: Love it or leave it? A new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26:98–117, 2009.

[13] R. Gschwind and M. Baumgartner. The digital Paul Klee - a case study. In *Proceedings of Archiving 2005*, pages 96–98. IS&T, Springfield, VA, 2005.

[14] I. Lissner and P. Urban. Toward a Unified Color Space for Perception-Based Image Processing. *IEEE Transactions on Image Processing*, 21(3):1153–1168, 2012.

[15] M. Kleiner, D.H. Brainard, and D Pelli. What's new in Psychtoolbox 3? *Perception (ECVP 2007 conference abstracts)*, 2007.

[16] Sanjoy Dasgupta and Daniel Hsu. Hierarchical sampling for active learning. In *Proceedings of the 25th International Conference on Machine Learning*, pages 208–215, 2008.

[17] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[18] Ch. Diehl and G. Cauwenberghs. SVM incremental learning, adaptation and optimization. In *Proceedings of the 2003 International Joint Conference on Neural Networks*, pages 2685–2690, 2003.

[19] M. Scheller Lichtenauer, I. Sprow, and P. Zolliker. Choice based experiments in multiple dimensions. *Color Research and Application*, accepted 2011.