Conjoint Analysis of Parametrized Gamut Mapping Algorithms

Zofia Barańczuk, Iris Sprow and Peter Zolliker, Laboratory for Media Technology; Swiss Federal Laboratory for Materials Testing and Research (EMPA), Dübendorf, Switzerland
Joachim Giesen, Friedrich-Schiller-Universität Jena, Germany

Abstract

We show that conjoint analysis, a popular multi-attribute preference assessment technique used in market research, is a well suited tool to simultaneously evaluate a multitude of gamut mapping algorithms with a psycho-visual testing load not much higher than in conventional psycho-visual tests of gamut mapping algorithms. The gamut mapping algorithms that we test using conjoint analysis are derived from a master algorithm by choosing different parameter settings. Simultaneously we can also test the influence of additional parameters like gamut size on the perceived quality of a mapping. Conjoint analysis allows us to quantify the contribution of every single parameter value to the perceived quality.

Introduction

Rendering of a color image in the presence of device limitations, also called gamut mapping, is a fundamental problem in digital color reproduction. Despite being a classical topic, for an overview see Morovic [1], gamut mapping is still an active area of research. Lately, research on gamut mapping algorithms (GMAs) has focused on image dependence [2, 3] and spatial mapping algorithms [4, 5, 6]. A very important part in the development of GMAs is their evaluation. Here human perception is the ultimate judge that determines which of the different competing algorithms is the most effective. Psycho-metrical scaling is a common method to measure image quality and image differences [7]. The quality of GMAs is typically measured with psycho-visual tests that involve paired comparisons. In a paired comparison a test person is shown an original image and two images obtained from different mapping algorithms. The test person has to identify the mapped image perceived to better represent the original. In order to improve the quality and comparability among studies, the technical committee of CIE published guidelines [8] on how to conduct psycho-visual tests assessing the quality of GMAs.

Here we use psycho-visual tests not only to compare a few final GMAs but already in the development stage of mapping algorithms. Our approach builds on the insight that gamut mapping can be seen as a highly parametrized problem. There are many, sometimes competing parameters relevant for gamut mapping: first of all the preservation of hue, lightness and saturation. Important parameters for image dependent algorithms are spatial image information and parameters like local contrast and smoothness gradients. Also, when realizing a GMA we have a choice of working color space, mapping direction, compression type. We use psycho-visual tests—paired comparison—to determine an optimal parameter setting. The data elicitation phase of our test is the same as in traditional psycho-visual tests used to compare different GMAs. In particular, the number of paired comparisons per test person is not larger, neither is the number of test persons significantly larger, whereas the potential number of mapping algorithms that can be compared is much larger. The difference to traditional psycho-visual tests used to compare GMAs is in the way how we analyze the elicited data. We are using conjoint analysis that essentially fits a linear model [9] to the data by assigning a part-worth value to every parameter level. The value of a parameter setting—besides the algorithm's parameters this can include additional parameters like gamut size—is then the sum of the part-worth values of the parameter levels used. The number of potential parameter settings that can be compared using conjoint analysis is determined by the number of levels tested for each parameter, i.e., it is the product of these numbers which can be quite large and easily exceed 1000.

We should point out that we are not the first who systematically include observer experiments in the development of GMAs, see for example the work done by Kang et al. [10]. Multivariate analysis techniques also have been used in image processing to gauge the importance of parameters [11].

This paper is organized as follows. In the next section, "Mapping Algorithms", we describe the parameters that we have studied and evaluated for gamut mapping. Section "Conjoint Analysis" reviews a conjoint analysis technique which was developed in [9] by extending Thurstone's law of comparative judgment to the multi-attribute case. In Section "Results" we present and discuss two user studies that we conducted to evaluate the parametrized GMA. We conclude the paper with a discussion of our results.

Mapping Algorithms

We consider finding a good GMA as a parameter optimization problem, i.e., we consider one master algorithm with free parameters for which we want to determine optimal values from psycho-visual tests. The master algorithm is quite simple, it maps any color point in the source gamut along a line segment connecting the color point and a focal point into the destination gamut. We consider also additional parameters that are not parameters of our master algorithm, but whose variation may affect the perceived quality of the mapping. We are especially interested in how the shape of the destination gamut affects the quality of the mapping. Hence we considered the following additional parameters: the size of the destination gamut as well as small shifts and rotations of the destination gamut. Shifts and rotations of destination gamuts turned out to be not so important parameters. Thus in a second user study we replaced them by two other parameters, namely shift of the focal point (Color/Density shifts) and hue rotations. In the following we summarize all the parameters that we have studied. Note that we always used sRGB as source gamut, i.e., we did not consider the source gamut as a parameter.

Compression. The compression parameter describes how the mapping along the line segment is done. We studied four different strategies: linear compression (lin), clipping (clip) and two sigmoidal compression algorithms (sig1 and sig2). The sigmoidal compression strategy maps a color point towards the focal point by a scale factor $0 < \beta \leq 1$ calculated as follows

$$\beta = \alpha \cdot D \cdot \tanh\left(\frac{S}{D} \cdot \tanh^{-1}\left(\frac{X}{S}\right)\right) + (1-\alpha) \cdot \frac{X \cdot D}{S},$$

where X is the distance of the focal point to the color point that needs to be mapped, S is the distance of the focal point to the source gamut boundary, and D is the distance of the focal point to the destination gamut boundary. The two sigmoidal compression strategies differ only by the choice of α which is 0.5 for sig1 and 0.8 for sig2.

Detail enhancement. As Zolliker [4] has shown, enhancing details can improve the quality of the mapped image essentially. Detail enhancement is a procedure independent of the master algorithm. Actually, it is technically more involved than the master algorithm itself. But we can interpret it as a parameter of the master algorithm in the sense that we can apply detail enhancement in varying degrees to the results obtained from the master algorithm. We use the detail enhancement method described in [4] with different levels of enhancement, namely 0.5, 1, and 1.5, respectively. The levels of the detail enhancement parameter are E1, E2, E3, and N which means that no detail enhancement was applied.

Hue preservation. Hue preservation is considered an important aspect in gamut mapping. We test its importance in two ways: one is to use two different working color spaces CIE-LAB [12] and IPT [13]. CIE-LAB is known to preserve hue not accurately, especially in the blue region, whereas IPT was designed to preserve hue. The other way is to add specific hue shifts to the image. Thus the first two levels of this parameter describe the working color space used for our master algorithm (called LAB, IPT). Two more parameter levels (called IPT+ and IPT-) add $+0.1$ radians or -0.1 radians, respectively to the image. Here IPT was used as working color space.

Gamut size. To gauge the importance of the destination gamut we also tested a parameter that describes the size of the destination gamut. This is actually not a free parameter of our master algorithm, but we decided to include it, because it allows us to estimate the relative importance of the destination device gamut size compared to the free parameters of the master algorithm. We tested four different destination gamuts. The smallest was ISONewspaper, the largest ISOCoated. The two remaining gamuts were created from ISONewspaper and ISOCoated gamut, as weighted average with weights $\frac{2}{3}$ and $\frac{1}{3}$, and $\frac{1}{3}$ and $\frac{2}{3}$, respectively. We refer to the levels of the gamut parameter as sml, med1, med2, and lrg, respectively.

Color/Density shifts. Another free parameter of our master algorithm is the choice of focal point. The idea is to produce well defined color and density shifts in the mapped image by varying the focal point. A natural choice for the focal point is close to the mid point of the gray axis in the destination gamut. In our case we choose the mid point of our smallest gamut, i.e. $(59,0,0)$ in LAB space. We also used focal points shifted along the gray axis by vectors $(5,0,0)$ and $(-5,0,0)$ and shifted in the color plane by $(0,3,0)$, $(0,-1.5,1.5\sqrt{3})$,$(0,-1.5,-1.5\sqrt{3})$. We refer to the six levels of this parameter in the order we have described them as D, D+, D-, a+, a-b+, and a-b-, respectively.

Gamut shifts. This parameter describes a shift of the destination gamut. We considered six levels with shifts identical to those of the 'Color/Density shift' parameter. We refer to the levels of this parameter in the order that we have described them as G-D, G-D+, G-D-, G-a+, G-a-b+, and G-a-b-, respectively.

Gamut rotations. Another parameter that we considered is a rotation of the destination gamut. The first two levels are identical to the levels of the 'Hue Preservation' parameter, namely the use of CIE-Lab and IPT color space. Two additional levels describe rotations of the destination gamut in the IPT in IPT space by -0.1, 0, $+0.1$ radians. We refer to the levels of this parameter as $G-LAB$,$G-IPT^{-}$, $G-IPT$, $G-IPT+$.

Conjoint Analysis

Conjoint analysis comprises a plenitude of techniques to disaggregate measurements on a parametrized domain. We call a domain parametrized, if it is given as a Cartesian product $A = A_1 \times \ldots \times A_m$ of parameter sets $A_1, \ldots, A_m$. Every element of A is a vector $(a_{1i_1}, a_{2i_2}, \ldots, a_{mi_m})$, where $a_{ki} \in A_k$. The elements a_{ki} are called parameter levels. One goal of conjoint analysis is to determine how much every parameter level contributed to the observed outcome of a measurement—this is called the part-worth of the parameter level. One particular form of measurements that we want to consider here are paired comparisons. In a paired comparison from two given elements $a, b \in A$ one is chosen to be better than (or preferred to) the other.

In [9] a conjoint analysis technique was developed as an extension of Thurstone's method [14] of comparative judgment to the multi-parameter (conjoint) case. The extension entails to apply Thurstone's method for each parameter to compute an initial set of part-worth values. Then rescaling the values makes the scales of the different parameters comparable. The overall value of an object in A (in our case an incarnation of our master GMA) is obtained by summing up the re-scaled part-worth values of the parameter levels present in the object.

Thurstone's Method

Let us briefly recapitulate Thurstone's method [14] of comparative judgment that is often applied to compare GMAs. The method is described in detail in [7].

Given is a finite set of stimuli, e.g. GMAs. According to Thurstone's method, the frequency of preferring stimulus i to stimulus j ($F_{i \succ j}$) is an indirect measure of the distance in quality of the stimuli i and j. This distance is the difference in the expected (quality) values of i and j. We assume that quality values S_i of the stimuli are uncorrelated, normally distributed random variables with expectation μ_i and equal variances, i.e. $\sigma_i^2 = \sigma^2$. These assumptions will be supported by the test that we describe later. By the properties of normal distributions, the differences $\Delta S_{ij} = S_i - S_j$ are also normally distributed with expected values $\mu_i - \mu_j$ and variances $2\sigma^2$, so $\Delta S_{ij} = \Phi^{-1}(P(i \succ j))$, where Φ is the cumulative distribution function of a normal distribution with variance $\sigma' = 2\sigma^2$. In this case the quality value difference $\mu_i - \mu_j$ equals $\sqrt{2}\sigma\Delta S_{ij}$, since the variance of the distribution has doubled. We estimate $P(\Delta S_{ij} > 0)$ by the observed $F_{i \succ j}$. Having a matrix of probabilities we can calculate $\mu_i - \mu_j$

for all $i, j = 1, \ldots, n$ where $i \neq j$. We can assume without loss of generality that the average of the quality values $\overline{\mu}$ is 0. Hence we compute μ_i for $i = 1, \ldots, n$.

$$\frac{1}{n}\sum_{i=1}^{n}(\mu_i - \mu_j) = \mu_i - \frac{1}{n}\sum_{j=1}^{n}\mu_j = \mu_i - \overline{\mu} = \mu_i$$

We can see, that σ does not change the ratio of quality value differences. Hence a natural choice of σ is $\sigma = 1$.

In some cases we do not have a full frequency matrix since not all stimuli have been compared with each other. Hence we cannot apply Thurstone's method directly. We fix this by first applying the method to pairs of stimuli that have been compared, and then we use linear regression to expand our results to all pairs of stimuli.

The multi-parameter case

We now assume that our stimuli come from a structured set, i.e., every stimulus is an element of a set $A = A_1 \times \ldots \times A_m$, where the A_k are finite parameter domains that describe the stimuli. Preference of one stimulus to another is measured by paired comparison. To apply Thurstone's method directly to A we need a large number of paired comparisons since the set A typically is large, i.e., $|A| = |A_1| \cdot \ldots \cdot |A_m|$. Furthermore, we aim for more information here, namely, if possible we want to measure the contribution of each parameter level to the observed overall quality of a stimulus. Therefore we use a decompositional approach. First, we compute quality values for the levels of each parameter $A_1, \ldots, A_m$ using Thurstone's method with a probability matrix created as follows: for each parameter A_k, if $(a_1, \ldots, a_m) \in A$ was preferred to $(b_1, \ldots, b_m) \in A$, then we interpret this as a_k was preferred to b_k if $a_k \neq b_k$.

Note that when applying Thurstone's method directly we use $\sigma = 1$. Now let $s_{k_1}, \ldots s_{k_{n_k}}$ be the quality values computed using Thurstone's method for every stimuli A_k with levels $a_{k_1}, \ldots a_{k_{n_k}}$. To get the quality value of a stimulus we sum up all the quality values (part-worth) of the parameter levels present in the stimulus, i.e., we make the assumption of a linear model. But in order to compare the quality values for different parameters we have to normalize them. We derive a normalization procedure from the following assumption.

Assumption. For any parameter A_k the quality values $s_{k_1}, \ldots s_{k_{n_k}}$ are normally distributed with variance σ_{k1}^2 and expected value 0 drawn from another normal distribution with expected value 0 and variance σ_{k2}^2. Hence, scale values for the levels of parameter A_k are drawn from the normal distribution N_k with variance $\sigma_{k1}^2 + \sigma_{k2}^2$ and expected value 0 (as the convolution of two normal distributions with expected value 0 and variances σ_{k1}^2 and σ_{k2}^2 respectively).

The value σ_{k1}^2 is equal for all s_{k_j} and will be chosen such that comparable quality values for different parameters, i.e., the values $\sigma_{k1}s_{k_j}$ will be comparable. As we compute quality values from paired comparison on stimuli level, the stimuli quality values by our assumption all are drawn from the distributions $N_k, k = 1 \ldots m$. Hence all these distributions should be the same, i.e., the value $\sigma_{k1} + \sigma_{k2}$ is independent of $k = 1, \ldots, m$. Without loss of generality we can assume that $\sigma_{k1} + \sigma_{k2} = 1$ for $k = 1, \ldots, m$. In the following we fix the parameter k and drop it from the index. We can estimate σ_2 from the re-scaled observed quality values $\sigma_1 \Delta S_{ij}$ by using an estimator of the standard deviation, i.e.,

$$\sigma_2 = \sigma_1 \sqrt{\frac{\sum_{i=1}^{n_i}\sum_{j=1}^{n_j}\Delta S_{ij}{}^2 F_{i\succ j}}{2\sum_{i=1}^{n_i}\sum_{j=1}^{n_j}F_{i\succ j}}}$$

Given that $\overline{\Delta S_{ij}} = 0$ and because of $\sigma_1^2 + \sigma_2^2 = 1$, we get:

$$\sigma_1 = \frac{1}{\sqrt{1 + \frac{\sum_{i=1}^{n_i}\sum_{j=1}^{n_j}\Delta S_{ij}{}^2 F_{i\succ j}}{2\sum_{i=1}^{n_i}\sum_{j=1}^{n_j}F_{i\succ j}}}}$$

Now for the fixed parameter k we re-scale the values $s_{k_1}, \ldots, s_{k_{n_k}}$ computed by Thurstones method by the estimated value of $\sigma_{1k}(= \sigma_1)$ to normalize them. The normalized quality values of the parameter levels are our part-worths that we assume to contribute linearly to the quality of a stimulus, i.e., the quality value of a stimulus $(a_{1_{j_1}}, \ldots a_{m_{j_m}}) \in A$ is $\sum_{k=1}^{m}\sigma_{k1}s_{k_j}$ which is the sum of the part-worths of the parameter levels present in the stimulus.

Error analysis

We estimated errors of our part-worths in the following ways.

Linear regression. Note that we can treat any comparison of stimulus i and stimulus j as an independent Bernoulli trial with probability of success (i is preferred to j) equal to p. We can estimate p by $F_{i\succ j}$, as $F_{i\succ j}$ converges to p when the number of trials goes to infinity. For a finite number of trials as in our case the standard deviation of a Bernoulli trial is estimated as

$$\sigma = \sqrt{\frac{F_{i\succ j}(1 - F_{1\succ j})}{m_{ij}}}$$

where m_{ij} is the number of comparisons of items i and j. To compute the errors of the part-worths from the frequencies, we use error propagation.

Error estimation. We also computed errors experimentally. To compute the experimental error we divided the paired comparison randomly into two groups. For each group we computed the part-worths individually. We repeated this process several times and averaged the absolute difference of the results for every part-worth, i.e., every level of every parameter. This average provides us with an experimental error. Similarly we computed experimental errors by dividing images into groups and by dividing test persons into groups.

Data Assessment

We conducted two studies which we refer to as "Symposium test" and "EMPA test". The first test took place at the 2008 Color Management Symposium in Munich, and the second test was conducted at our institute (EMPA). In the second test we replaced the additional parameters destination 'Gamut shift' and 'Gamut rotations' by the master algorithm parameter 'Color/Density shifts' and 'Hue preservation', otherwise the set-up of the tests was the same.

Test procedure. For every paired comparison a participant of our test was shown an original image and two images mapped with different parameter settings on a LCD screen. The original image was presented in the upper half of the screen and the two

Figure 1. *Test application: user interface.*

mappings below the original side by side, see Figure 1. The two mappings were chosen at random from our parameter space. In the 'Symposium test' we used the constraint, that levels of 'Gamut size' in compared images can differ by only one consecutive level (in the natural order) since larger differences in gamut size essentially determine the choice. The test person who participated in our tests had to choose the mapped image that reproduces the original better. For their choices the test persons used a mouse to click on the corresponding image. If no difference could be seen, the original had to be selected in order to avoid a forced choice.

Monitor. For the test we used 24" Eizo CG 241W and 22" Eizo CG 220 LCD monitors calibrated to show sRGB correctly with a white point set to 6500 Kelvin. The ambient illumination measured in the middle of the switched off monitor was at 40 lx. Monitor flaps around the screen prevented flare. The monitor's background was set to a neutral gray.

Test Images. The image set included the obligatory "Ski" image that is specified by the CIE guidelines. To average out a judgment bias due to image content, a wide range of scenes, including 97 different images, was used in the experiment. In addition to a set of ISO test images, they were taken from royalty free libraries as well as from private stock . For each image 1536 mapping combinations were possible, resulting from the 5 parameters with a total of 22 levels. All images had a resolution of 400×600 pixels or 400×400 pixels on the 96 dpi screen which resulted in 8.5×12.5 cm or 8.5×8.5 cm sized images on screen.

Test persons. Test persons for the first test were recruited from participants of the 2008 Color Management Symposium in Munich who were mostly color experts and participated voluntarily. Each test person had passed the Ishihara test for color deficiency. We had 70 test persons participating in our study who each did 50 paired comparisons, resulting in a total of 3500 comparisons. The participants of the second test were mostly EMPA employees. Each test person had passed the Ishihara test for color deficiency. We had 13 test persons participating in our study who each did between 50 and 200 paired comparisons, resulting in a total of 2100 comparisons.

Results

In a first step we computed part-worths for all the different parameters individually. To ensure, that the obtained part-worths are meaningful, we tested the following two assumptions which underlie our computations:

(1) The scale values for all levels of a given parameter are uncorrelated, normally distributed and have the same variance.
(2) Linear Model, i.e, the quality value for a parametrized stimulus—in our case an image—can be obtained by summing up the part-worths of the parameter levels present in the stimulus. This essentially means that the part-worths for the different parameters are uncorrelated.

Mosteller's test. We used Mosteller's test to test the assumption on the parameter level that the part-worths are uncorrelated, normally distributed variables with equal variances. A description of Mosteller's test can be found in [7] or [15]. Results are presented in Table 1. All parameters passed the test at a significance level $\alpha = 0.1$.

	Mosteller	$\chi^2, \alpha = 0.1$
Symposium test		
Gamut Rotation	2.4	6.3
Gamut Shift	11.2	16.0
Compression	7.7	6.3
Detail Enhancement	2.8	6.3
Empa test		
Hue preservation	3.6	6.3
Compression	2.3	6.3
Color/Density	14.7	16.0
Gamut Size	7.3	6.3
Detail Enhancement	5.6	6.3

Table 1. Mosteller's test for parameters compared to χ^2 with significance level $\alpha = 0.1$

Linear Model. We tested the linear model assumption by testing preferential independence of pairs of parameters. Let A_1 and A_2 be two parameters, let $C = A_1 \times A_2$ be their Cartesian product, and let $c_1, \ldots, c_k$ be the levels of C. We compute part-worths for $c_1, \ldots, c_k$ in two ways. First, we compute part-worths as described in Section "Conjoint Analysis" for all levels of the parameters A_1 and A_2. For every level $c_i = (a_{i1}, a_{i2})$ with $a_{i1} \in A_1$ and $a_{i2} \in A_2$ we add scale values for a_{i1} and a_{i2}, getting results $s_1, \ldots s_k$. Second, we apply Thurstone's method directly to the combined parameter C. We rescale the scale values as described in Section "Conjoint Analysis". We denote the results as $s'_1, \ldots s'_k$. If the parameters are additive (preferentially independent), then we should get $s_i \approx s'_i$. We test a null hypothesis that $s_i = s'_i$ for all $i = 0, \ldots k$ by using a χ^2 test with the following test statistic:

$$\chi^2 = \sum_{i=1}^{k} \frac{(s_i - s'_i)^2}{\sigma_i^2 + \sigma_i'^2},$$

where σ_i and σ'_i are error estimations computed by linear regression from observed frequencies. The null hypothesis postulates that the test statistic is approximately χ^2 distributed with $k-1$ levels of freedom.

In both tests we observe no significant correlation between any parameters, except for the combination 'Gamut-Size'—'Density/Color' and 'Gamut-Size'—'Gamut-Shift', where a small correlation was found. However more data is needed to allow a more detailed conclusion.

Discussion

Here we demonstrate how the computed part-worths can be used to answer questions like: What is the relative importance of the different parameters? Which levels of the parameters are most preferred? Which preferential dependencies exist among the parameters?

Importance of parameters. Importance of parameters describes how much each parameter contributes to the quality value on the stimulus level, i.e., the combination of all parameters. Two methods are often used to calculate the importance of parameters. The first one considers the largest difference between the part-worths within the parameters. The second one uses the standard deviation of the part-worths within the parameters. Note that for both methods the computed importance depends on the levels chosen for the parameters, e.g., if we choose levels for a parameter that hardly can be distinguished, then the importance of this parameter will be low, while it could be high for a different choice of levels. Here we applied the second method. The importance of the different parameters is shown in Table 2. In this table we also present the standard deviation $\sigma_{\Delta E}$ of ΔE (distance of the transformed image and the original, averaged over the images). We calculated ΔE for each parameter taking default levels of other parameters (med2, sig2, D0/G-D0, IPT/G-IPT, loc1) and computing the difference between the transformed image and the original one. Note that in general the importance of parameters correlates with the average difference ΔE, with exception of the 'Details Enhancement' parameter, which shows the smallest ΔE despite of its medium importance. This is not surprising, as local contrast conservation can not be measured by a global color distance measure such as ΔE.

Rank	Parameter	Importance	$\sigma_{\Delta E}$
	Symposium test		
1	Gamut Size	0.572	3.36
2	Compression	0.187	1.46
3	Detail Enhancement	0.130	0.23
4	Gamut shift	0.074	0.45
5	Gamut rotation	0.037	0.21
	EMPA test		
1	Gamut Size	0.473	3.36
2	Color Density	0.161	2.12
3	Detail Enhancement	0.150	0.23
4	Compression	0.131	1.46
5	Hue preservation	0.085	0.61

Table 2. Importance of parameters (scaled such that sum is 1)

The gamut size is the most important but not the only deciding parameter. In Figure 3 we can see, that there are no large gaps between successive quality values for all the 1536 parameter settings that we have tested. This means that preferred levels of other parameters can compensate for a samller gamut. Note that a Gaussian inverse cumulative distribution is a good approximation of the shown experimental curve in agreement with our model assumptions.

Most preferred levels. Figure 2 presents part-worths for all parameter levels. Preferences concerning the gamut size are as expected, the larger the gamut, the higher is the perceived quality of the image. Clipping emerges as the best method of compression. Its disadvantage of loss of details in saturated colors is compensated, because part of the details are reconstructed by the 'Details enhancement' parameter for all but the N-level. This result confirms, that saturation is an important factor in gamut mapping. This is also supported by the high importance of the gamut size parameter. As expected linear compression performs worst. The parameter 'Detail enhancement' shows a strong preference for higher enhancing factors. The gain between 1 and 1.5 is smaller than those between 0 and 0.5 or 0.5 and 1. This indicates that there is an upper limit to detail enhancement, but the maximum may be beyond the 1.5 setting. For the 'Color/Density shifts' we would expect a setting on the neutral axis to be most preferred. This is true as all images with color shifts perform worse than the two preferred settings D0 and D-. This indicates that the D- setting is closest to ideal setting of the focal point.

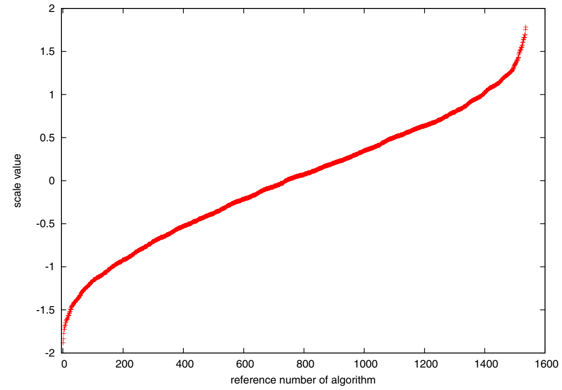

Figure 3. Sorted scale values for algorithms, the Symposium test

The part-worths are similar for the parameters that were tested in both tests. In the EMPA test 'Details enhancement' seems to be slightly more important compared to compression. Also, in the EMPA test the parameter Color/Density, not present in the Symposium test was quite important, which means that the other parameters contributed less to the general result.

Parameter inter-dependencies. Testing the linear model showed that there is no need to use a model with combined attributes levels.

Error analysis. We did not notice a large difference between the experimental error computed by randomly dividing the paired comparisons and the error calculated by linear regression. The experimental error computed by randomly dividing the images however is notably larger. This suggests that it is worthwhile to develop algorithms based on individual image properties. In this experiment we had not enough data to conduct a more detailed analysis in that direction.

Conclusions

We showed that conjoint analysis can be a useful and efficient method to gauge the importance of gamut mapping parameters for the perceived visual image quality. Our research confirmed a few known results about level preferences, e.g., large gamuts preferred over small ones. The real strength of the method however is, that such a multi-parameter study allows to compare

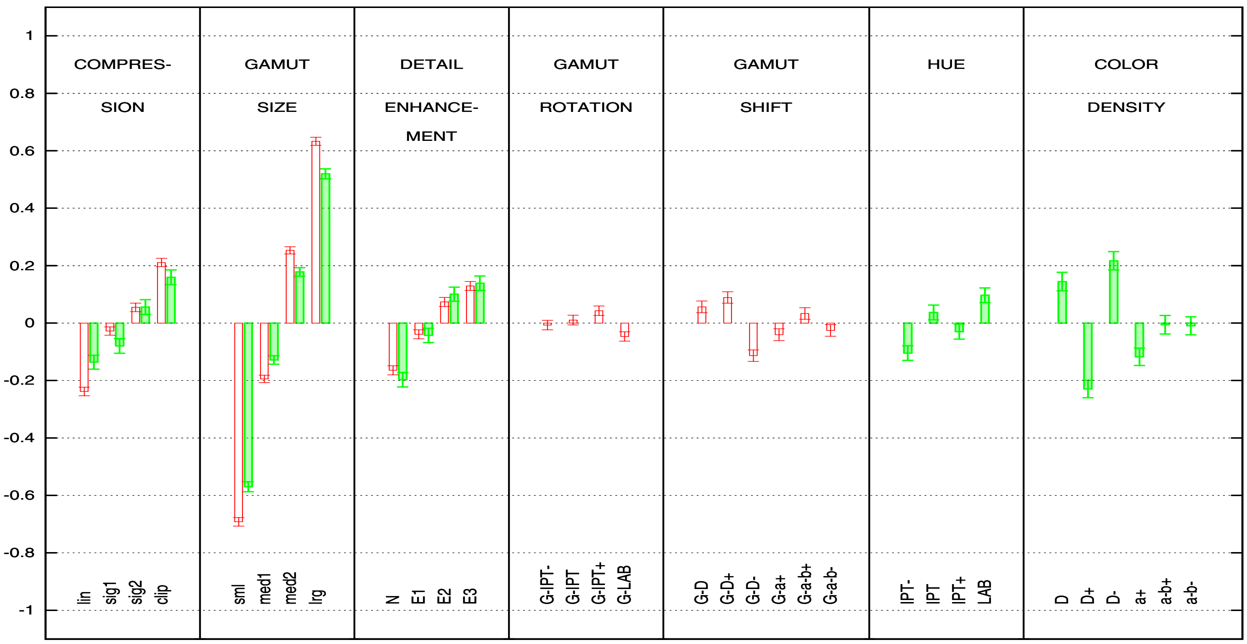

Figure 2. *Part-worths for all parameter levels. The red graph (not filled) shows results for the Symposium test, the green (filled) one for the EMPA test.*

the importance of different levels across different parameters. For example we consider noteworthy that the working color space is a surprisingly insignificant parameter and parameters like compression or detail enhancement are more important. This probably means, that it is more interesting to focus research on the latter parameters than improving the working color space. Another important finding is that in some cases a good selection of GMA parameter levels can compensate for limited gamut size.

Acknowledgments

This research was supported by the Hasler Stiftung Proj. # 2200. We thank Fogra for allowing us to collect data at the Color Management Symposium, EIZO for providing us monitors, and everyone participating in our studies.

References

[1] J. Morovic. *Colour Gamut Mapping*. WileyBlackwell, ISBN 0470030321, 2008.

[2] H. Kotera and R. Saito. Compact description of 3-d image gamut by r-image method. *Journal of Electronic Imaging*, 12(4):660–668, October 2003.

[3] J. Giesen, E. Schuberth, K. Simon, and P. Zolliker. Image-dependent gamut mapping as optimization problem. *IEEE Transactions on Image Processing*, 16(10):2401–2410, October 2007.

[4] P. Zolliker and K. Simon. Retaining local image information in gamut mapping algorithms. *IEEE Transactions on Image Processing*, 16(3):664–672, March 2007.

[5] N. Bonnier, F. Schmitt, M. Hull, and Leynadier C. Spatial and color adaptive gamut mapping algorithms. In *Color Imaging X: Processing, Hardcopy and Applications*, volume 15, pages 267–272, Scottsdale, AR, 2007. IS&T/SID.

[6] I. Farup, C. Gatta, and A. Rizzi. A multiscale framework for spatial gamut mapping. *IEEE Transactions on Image Processing*, 16(10):2423–2435, October 2007.

[7] Peter G. Engeldrum. *Psychometric Scaling, A Toolkit for Imaging Systems Development*. Imcotek Press, Winchester MA, USA, 2000.

[8] Central Bureau of the CIE, Vienna. *CIE Publication 156: Guidelines for the Evaluation of Gamut Mapping Algorithms*, 2004.

[9] J. Giesen, K. Mueller, E. Schuberth, L. Wang, and P. Zolliker. Conjoint analysis to measure the perceived quality in volume rendering. volume 13, pages 1664–1671, November 2007.

[10] Byoung-Ho Kang, Maeng-Sub Cho, Jan Morovic, and Ronnier M. Luo. Gamut compression algorithm development on the basis of observer experimental data. In *Eighth Color Imaging Conference: Color Science, Systems, and Applications*, volume 8, pages 268–272. The Society for Imaging Science and Technology, November 2000.

[11] B. Keelan. *Handbook of Image Quality*. Marcel Dekker, Inc., 2002.

[12] Central Bureau of the CIE, Vienna. *CIE Publication 116: Industrial color difference evaluation*, 1995.

[13] Fritz Ebner and Mark D. Fairchild. Developement and testing of a color space (ipt) with improved hue uniformity. *Proceedings of 6th IS&T/SID Color Imaging Conference, Scottsdale AZ*, pages 8–13, 1998.

[14] L. Thurstone. A law of comparative judgement. In *Psychological Review*, pages 273–286, 1927.

[15] F. Mosteller. Remarks on the method of paired comparisons: III. a test of significance for paired comparisons when equal standard deviations and equal correlations are assumed. *Psychometrika*, 16:203, 1951.

Author Biography

Zofia Barańczuk received her bachelor's degree in computer science and master's degree in mathematics from the Warsaw University. She works now on her doctorate and is engaged in psycho-visual tests and gamut-mapping.