Lux junior 2023

16. Internationales Forum für den lichttechnischen Nachwuchs
Ilmenau, 23. – 25. Juni 2023
DOI (proceedings): 10.22032/dbt.55787

DOI: 10.22032/dbt.58614

Evaluating colour Preference by using Multidimensional Approaches

Mark Suppelt, Julian Klabes, Paul Myland, Prof. Tran Quoc Khanh

Technische Universität Darmstadt, Fachgebiet für Adaptive Lichttechnische Systeme und Visuelle Verarbeitung, Hochschulstraße 4a, 64289 Darmstadt

Abstract

Colour preference is a key factor in the design and evaluation of lighting systems, particularly with the emergence of multichannel LED systems which allow for greater control over the spectrum of light emitted and therefore the colour appearance of the illuminated objects. To more accurately and objectively measure colour preference, there has been a growing interest in the development of multidimensional evaluation algorithms that consider multiple dimensions of colour rendering, such as chroma and hue shift. The purpose of this study was to compare and evaluate the performance of different multidimensional evaluation algorithms for colour preference in lighting applications. Using computer-generated images of a coloured object displayed on a computer monitor under a fixed white point, we simulated the colour shifts of the object under different light sources and test subjects evaluated the results using a range of multidimensional methods. Our analysis revealed that there are significant differences in the performance of these algorithms, with some providing more accurate and reliable measures of colour preference than others. Considering all relevant criteria, genetic algorithms seem to provide the most promising approach, as they lead to a result quickly and reliably. These findings have important implications for the selection and use of multidimensional algorithms for evaluating colour preference in lighting, particularly in the context of multichannel LED systems, and can inform future research in this area.

Index Terms: Lighting, Psychophysics, Colour preference

1 Introduction

Modern lighting systems - often referred to as 'smart lighting' - increasingly offer the ability to control several individually, addressable, coloured LED channels to mix white lighting. Consequently, this mixing of multiple LEDs leads to differences in white point and colour appearance of illuminated objects. However, many people lack the equipment or experience to quickly set the lighting scene to a preferred state. The main objective of this work is to determine the most optimal approach to achieve one's preference. For this reason, several psychophysical models will be tested and compared against each other. These models should offer a quick and most importantly repeatable methodology to assess a lighting scene.

The focus lies in establishing ideal methods, which work with multiple parameters. To be more precise, an image will be modified to change its hue and chroma values. Hue refers to the degree in which the stimulus can be described by red, green, blue, or yellow and it is specified by an angle around the achromatic white point in a colour space. Chroma is closely related to the saturation of a colour. It is described by Mark D. Fairchild as the “colorfulness of a stimulus relative to the brightness of a stimulus that appears white under similar viewing conditions” [1]. While colours, which are less saturated, also appear darker, a change in chroma does not affect the perceived brightness of said colour. It is specified by a radius or distance to the achromatic white point in a colour space. Modifying the images digitally has the benefit that apart from being simpler and more consistent, it allows for much finer control. Four methods and one control method, which are able to alter these values, are tested. The methods will be referred to as *Staircase*, *1D-Gauss*, *2D-Gauss*, *Genetic* as well as *Reference* respectively and are thoroughly detailed in section 2.2.

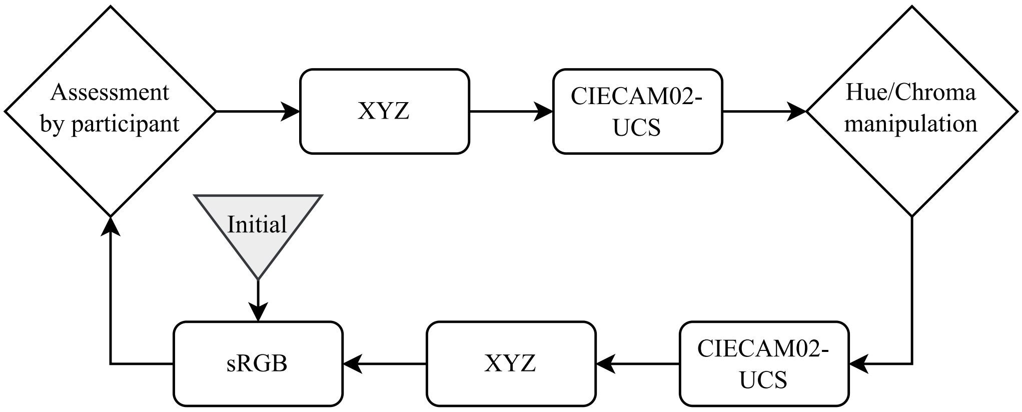

Figure 1-1 A typical workflow for the transformation and manipulation of the image

Most of the image manipulation is done using the python package ‘LuxPy’ [2]. A sRGB image, specifically of a red rose, is transformed into the CIECAM02-UCS format, where subsequently the Jab Coordinates for each pixel can be transformed to

$$Chroma = \sqrt{a^2 + b^2}$$
$$Hue = \arctan\left(\frac{b}{a}\right).$$

In the following sections the chroma and hue values of each pixel will be modified to change the appearance of the reference image. The Values are modified by either multiplication or addition/subtraction.

$$Chroma_{\text{new}} = Chroma \cdot Mod_{\text{Chroma}}$$
$$Hue_{\text{new}} = Hue + Mod_{\text{Hue}}$$

Mod_{Chroma} as well as Mod_{Hue} will be referred to as their corresponding modifiers. A typical workflow of the manipulation of images during the experiment can be seen in Figure 1-1. The image of a red rose (visible in Figure 2-1) has specifically been chosen since red is especially common in skin tones. For this reason, it has great influence on the appeal of a scene. [3]

1.1 Related work

Using human perception as a metric, mainly utilized in the field of psychophysics or psychometrics, allows for a range of methods to be used. Typical methods, like the method of limits, of constant stimuli or of adjustment, focus on detecting the thresholds of Perception. They are often used to detect the minimal auditory or visual thresholds but may suffer from anticipation bias and can be time consuming. For this reason usually adaptive psychophysical methods, most notably the Staircase method, are realized [4]. The Staircase method tries to approach a threshold value by incrementally stepping towards it and changing direction if it overshoots the threshold. Moreover, more complex methods like the PEST or Maximum-Likelihood-Methods do exist [5]. These improve the Staircase method by either adjusting the step size after each change in direction or carrying over information between trial runs. Additionally, H. Levitt suggests “Transformed Up-Down Methods” [6], which try to account for stimuli adaptation by e.g. requiring several negative responses before changing direction. However, most of these algorithms only optimize for one parameter at a time, while this study tries to find the ideal value and resulting combination for at least two parameters at the same time.

Another relevant method is scoring a selection of objects. In a study by Khanh *et al.* [7] a number of objects are compared with each other under different lighting scenarios. The participants are required to score them on a range of 0-100 for each different light source. This has the benefit of being able to directly compare different results or algorithms but suffers from effects like adaptation. One way to optimise the scene by using those scores is a genetic algorithm, which has already been comparably used by Newsham, G.R *et al.* [8]. The study tries to determine the preferred surface luminances in offices by creating randomized images and then optimising them similarly to an evolutionary process. A possible implementation is later discussed in section 2.2.3.

No forced-choice methods, described for example by Shelton, B. R. and Scarrow, I. [9], have been selected since this study's aim is to find a pre-emptive selection of useful algorithms for possible future studies. Forced-choice methods provide a selection of options and force the participant to pick one of them. Future studies are expected to be real-world applications of lighting scenes and not just virtual images. This would most likely rule out the possibility of showing a selection of scenes simultaneously and thus renders this type of algorithm inconsequential in this case.

2 Methods

A total of 28 participants have taken part in the experiment. All of which have been asked to provide general information like their age, sex, and if they have impaired vision, made apparent through the need for glasses or even colour blindness. Additionally, since a large pool of the participants have previous experience and technical ability in this field, they have been asked to disclose whether they consider themselves knowledgeable about this field. The results of these questions can be seen in Table 6-1. All the participants were required to find their preferred hue and chroma combination using several distinct methods, further detailed in section 2.2. Directly after the testing of each method, the participants have been asked to score the intuitiveness of said method in a range of 1–9 with 9 being the best possible score. Intuitiveness is meant to be a measure of ease of use or understandability.

2.1 Setup

The experiment takes place in an isolated room. The window as well as the entrance is covered by cloth to block any outside influences. The participant is placed in front of a monitor and has access to a controller. The monitor Color Edge CG277 (EIZO, Hakusan, Japan) has specifically been chosen for its colour accuracy and ability for self-calibration. The room is lit by two SkyPanel S60-C (ARRI, Munich, Germany) lights, placed behind the participant. Both have been modified to house an additional cyan channel for increased colour accuracy.

Figure 2-1 POV of the participant.

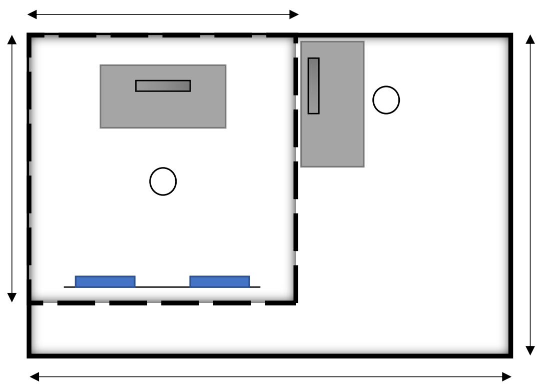

Figure 2-2 Schematic of the experimental setup.

All the surfaces in front of the participant as well as the monitor have been calibrated to be as close to the target visible in Table 6-3. Afterwards, the surfaces have also been measured by a spectroradiometer CS-2000 (Konica Minolta, Chiyoda, Japan) to validate them. The goal is to align the back- and foreground as much as possible to negate any other effects. The participants should ideally experience a homogeneous and reproducible state of adaptation. Special care has to be taken to reduce any outside influences since it could affect the models in a non-linear way [10]. A representative point of view can be seen in Figure 2-1. The participant is able to provide input via an Xbox Series One controller modified with an added keyboard. A schematic of the test room can be seen in Figure 2-2. The participant is in a secluded and covered area of the room as not to disturb the experiment, while the researcher is placed outside to control the experiment and to provide guidance. To guarantee no bleeding effect of external light, all other light sources in the room are shut off. The researcher reads from a prewritten script as not to influence the participants needlessly but is allowed to answer questions if they arise.

2.2 Algorithms

Several different methods have been selected to be compared against each other. All these methods, apart from *Reference*, are evaluated in a random order for each individual participant to negate any influence in between them. Before starting the experiment, a selection of the entire range images could be modified by is shown as a baseline or anchoring point. Figure 2-3 shows the entire process each participant must absolve.

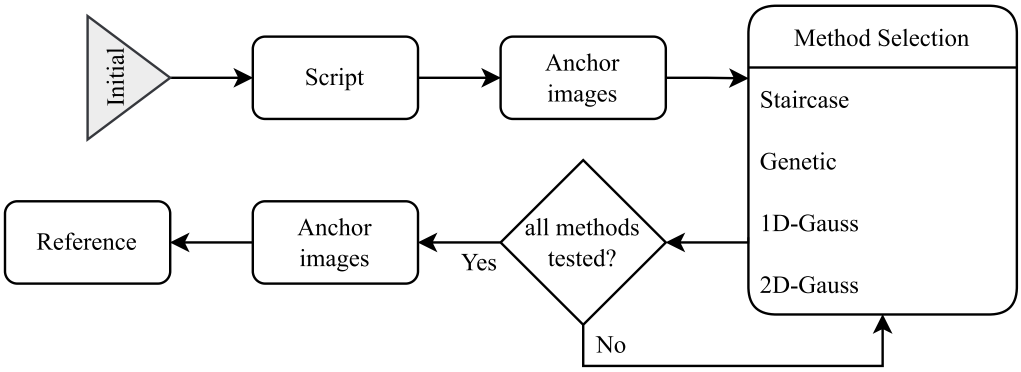

Figure 2-3 A flowchart of the procedure for each participant. Keep in mind that all methods apart from Reference are selected randomly.

2.2.1 Staircase

One of these methods is the so-called *Staircase* method. To be more precise the method closely resembles that of A. Pentland [11], which he calls “the best pest”, but it has been modified to allow for at least two parameters. The participant will be asked to compare a given image to the previous one. If it is better, the participant should press ‘1’, but if it is worse then ‘0’. The first image to be shown is the original image. Every time the participant approves of the new image the next one will have a modified hue or chroma value. The basic step size for Mod_{Chroma} and Mod_{Hue} are 8% and 4° respectively. However, each time the participants dislikes the new image the direction of change swaps and additionally, it halves in value i.e., instead of increasing the chroma value by 8% it is decreased 4% after the first direction change. This should limit the amount of overshooting the ideal value. After three direction changes the parameter is swapped, hence both the hue and chroma values are modified in an alternating manner. In total, both parameters are measured three times each, so that influences between the two parameters can be negated. However, the base image after each parameter swap is adjusted by the mean value of the direction changes. Let us suppose the participant pressed '0' for the chroma modifiers 1.2, 1.1 and 1.5, leading to a new base image for the next hue measurement with its chroma multiplied by $\frac{1}{3}(1.2 + 1.1 + 1.15) = 1.15$. The hue modifier is similarly carried over into the next measurement.

2.2.2 1D/2D-Gauss

The next two methods, *1D-Gauss* and *2D-Gauss* are similar. Both try to find the ideal hue and chroma combination by scoring several, randomly generated sample images and later fitting the results to a Gaussian function. The theory is, that the peak of the Gaussian function symbolizes the ideal or preferred value for the observed parameter. This has already been comparably used by S. Babilon *et al.* [12]. At first, a random image is generated by either multiplying the original chroma values with a random value or adding/subtracting from the original hue values. Both of the random values are selected from a uniform distribution and afterwards scored by the participant. The chroma multipliers range from 0.8-1.6, while the Hue values are modified in a range of ±13°. This procedure is used for two different methods. One fits to a one-dimensional

$$\mathrm{f(x) = f(y) = H + A \cdot exp}\left(-\frac{(\mathrm{x - x_0})^2}{(2\sigma^2)}\right)$$

and the other to a two-dimensional Gaussian function

$$\mathrm{f(x, y) = H + A \cdot exp}\left(-\left(\frac{(\mathrm{x - x_0})^2}{2\sigma_{\mathrm{X}}^2} + \frac{(\mathrm{y - y_0})^2}{2\sigma_{\mathrm{Y}}^2}\right)\right).$$

Please note that in case of more than two parameters, other, n-dimensional Gaussian functions also exist. One major difference between the two methods is that the two-dimensional function modifies the chroma (x) and hue (y) combination at the same time, whereas the one-dimensional function changes its modified parameter sequentially. After the image generation, the participant is then asked to score the image on a scale of 1-9, 9 being the best possible score. The score is used as an amplitude to which the Gaussian functions can be fitted to. The fitting is done via SciPy's [13] implementation of a Trust Region Reflective algorithm [14]. In total, the method for the one-dimensional Gauss takes two measurement series for both parameter with 20 samples each, while the two-dimensional Gauss takes 40 samples. This number was chosen to ensure a robust fit since, depending on the randomized samples or inconsistencies in scoring, the algorithm might not find an optimal solution if the sample size is too low. A possible fit for each algorithm can be seen in Figure 2-4 and Figure 2-5.

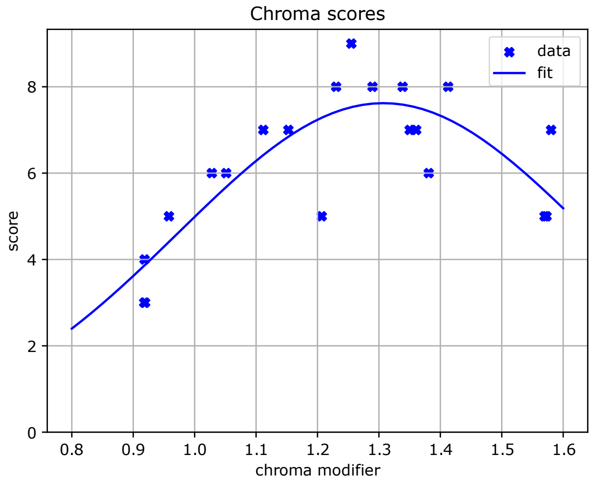

Figure 2-4 Sample fit of a 1D-Gauss for the chroma values. Fitting the hue values works likewise.

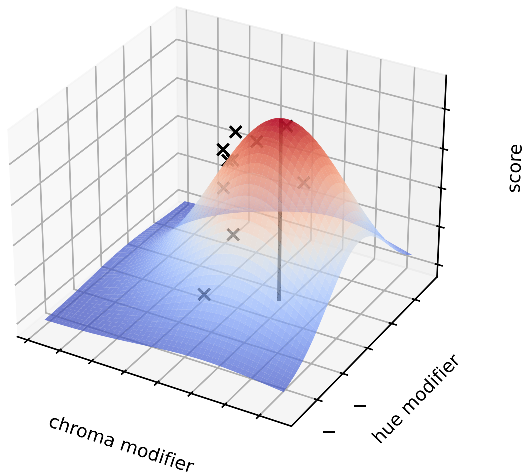

Figure 2-5 A sample fit for the 2D-algorithm. The ideal i.e., the maximum value is marked by the black line. For Illustrative reasons only ten sample are visible.

2.2.3 Genetic

Another proposed method is a *Genetic* algorithm, which has already been comparably used by Ashdown to find optimal luminaires [15] as well as by the previously mentioned paper of Newsham, G.R *et al.* to find preferred surface luminances in offices [8]. This method works by creating randomized images and then tries to identify the ideal composition by an evolutionary process. Specifically, this method has been implemented via pymoo's “GA: Genetic Algorithm” [16]. The population starts out with five images, which each possess differently modified hue and chroma values in the same range as for the Gaussian methods. The participant is then asked to score them in a range of 1-9 with higher scored images being more likely to survive and subsequently, passing their modifiers on to the next generation. The next generation inherits the parent's traits, nonetheless a small amount of mutation is also permitted. This process is repeated for five generations to find the ideal hue and chroma combination.

2.2.4 Reference

The last method is used as a control method and will always be the last one used, not to influence any of the following methods. It is similar to the *Staircase* method, but the user is able to independently adjust the image to their liking. The participants are able to change the hue and chroma values as well as the step size independently for as long as they need. The resulting image is used as a reference to compare all the other methods. This is needed since all participants are likely to have an individual ideal lighting scenario, which could conceal any differences made by the selected algorithm itself. Additionally, the anchoring image set, mentioned before, is shown right before this method as well to show the range of expected modification.

3 Results

For all the before mentioned methods, the generated hue and chroma combinations and if available the score, have been recorded. Additionally, after successfully concluding a method, the participants were asked to score it in a range of 1-9 on how intuitive they perceived said method. The results of which are displayed in Figure 3-1, while the individual scores of each participant can be seen in Table 6-1. Notably, the *Staircase* method not only presents the lowest average score, but also a significantly larger spread than the other methods. This indicates a polarizing nature of this method. Whereas all other methods are scored similarly positively. Next to it in Figure 3-2, the time needed to conclude each method is laid out. Again, the *Staircase* method is performing the worst, while the *Genetic* method is able to complete each run the most quickly.

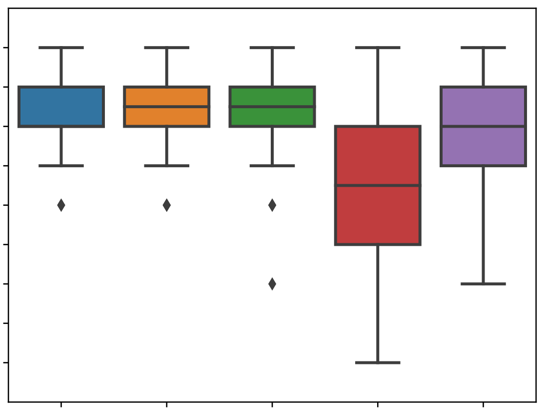

Figure 3-1 Scores in a range of 1-9 given by the participant on their perceived intuitiveness of the selected method.

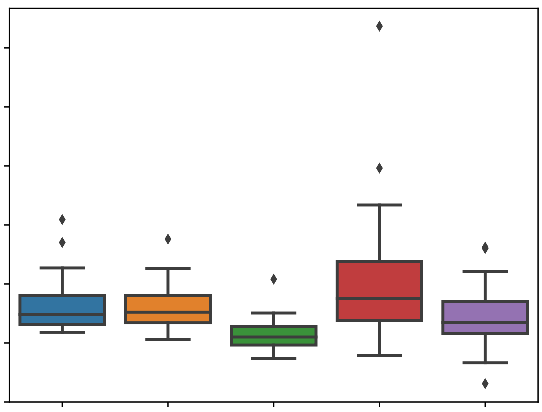

Figure 3-2 Time in minutes taken for the completion of each method.

Furthermore, the differences between the ideal hue and chroma values are of interest. For this reason, the *Reference* method is used as a control method. In theory, the participants should with limitless time and full control over the image, obtain their preferred scene. It is important to only use the relative modifiers.

$$\Delta Mod = Mod_{\mathrm{Method}} - Mod_{\mathrm{Reference}}$$

comparatively, since the preferred lighting scenes differ greatly between participants. For example, the preferred, absolute Mod_{Chroma} for *Reference* range from 0.83 to 1.97 (visible in Figure 6-1). However, *Reference* shows similarly to the other algorithms, that most people prefer an increase in chroma by roughly 15-40%. Also, most participants favoured an almost negligible increase in hue for the red rose, which is equivalent to a slight orange tint.

The relative modifiers are visible in Figure 3-3 and Figure 3-4. The large outliers, mainly visible for the *2D-Gauss*, can be explained by non-ideal fits. Also of note is the slight upwards shift of the *Staircase*'s chroma values.

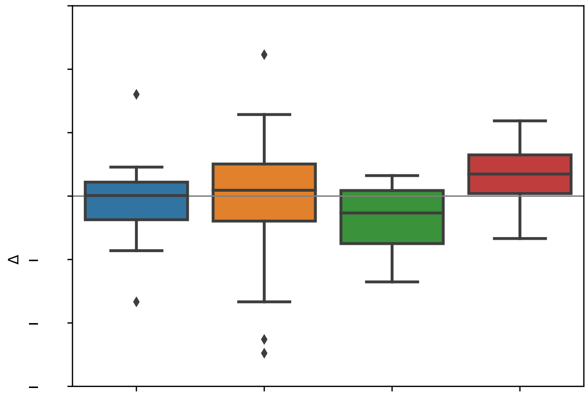

Figure 3-3 Difference of the chroma modifier between a method and Reference.

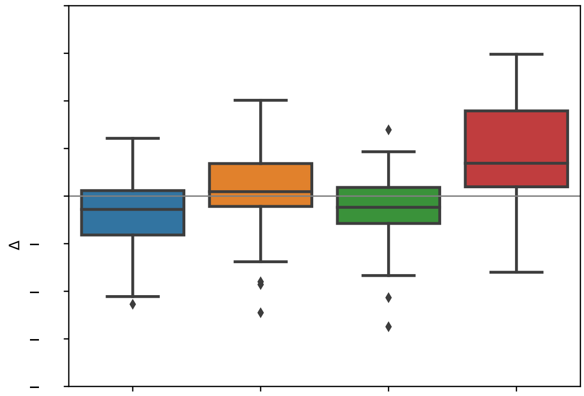

Figure 3-4 Difference of the hue modifier between a method and Reference.

The visible differences in the modifier values of each method have to be examined by statistical analysis. To proceed, one must first examine the distribution of the underlying data. For this reason, at first a Shapiro–Wilk test is performed to determine whether the data is normally distributed. With a significance level of 5% only the following methods are normally distributed: *Reference* the hue values of *Staircase* as well as the chroma values of *Genetic*. Consequently, neither an ANOVA nor t-test can be performed since they require a normally distributed sample set. On account of this and since the multiple measurements of each participant are related to each other, a Wilcoxon signed-rank test has been chosen. The Wilcoxon test is the non-parametric version of the paired t-test and checks whether two distributions, in this case all methods individually compared to *Reference*, differ significantly. The test concludes that, again with a significance level of 5%, only the *Staircase* modifiers as well as the chroma modifiers of *Genetic* differ significantly from the ones of *Reference*. The reasons for which will be further discussed in section 4. The precise statistical results can be seen in Table 6-2. [17] [18]

4 Discussion

During the comparison of all the methods the *Staircase* method stands out. Not only was it generally perceived as less intuitive than all the other methods, but it was also a lot more polarizing to the participants with a larger spread of scores. Besides that, it takes longer to complete, while also differing significantly from the control method *Reference*. A number of participants complained that they found it challenging to detect changes in the parameters since they forgot the previous image. As mentioned before in section 2.2, while a forced-choice algorithm i.e., showing a selection of images simultaneously would alleviate this problem, these kinds of algorithms have been ruled out with future studies with real world lighting applications in mind. Additionally, the implementation of the method itself might have influenced the outcome of the modifiers. All participants started out with the original image and by default, started by increasing the chroma levels. The resulting calculated modifier are visible in Figure 6-1. This fact as well as the difficulties in comparing the newly calculated images to the previous one, might have led participants to overshoot their preferred chroma level. The significant upwards shift in hue levels might also indicate the participants preference to approve the more recent images. To negate this effect, future studies should not only start at different points, but also switch the starting directions for each participant to limit any overshooting effects. In conclusion, while the *Staircase* method is quite popular and useful for many applications it might not be the best choice to optimise multiple parameters at once.

Additionally, looking closer at the participants themselves might reveal more information about the algorithms. For this reason, a correlation matrix (Figure 4-1) the information in Table 6-1 has been created. A Spearman correlation coefficient of 1 shows a strong monotonic relation whereas a Pearson coefficient displays a linear relationship. However, the information on the participants had to be normalized to be able to correlate it. In practise this means that ‘yes’ as well as ‘m’ have been converted to a one, while ‘no’ and ‘f’ have been converted to a zero. Participants with technical experience appear to dislike simpler methods like *1D-Gauss* and *2D-Gauss*, which require only minimal input by the user i.e., a rating. However, this effect seems to rescind with more complex methods like *Staircase* and *Reference*. Other factors such as age or sex might be skewed by the fact that most experienced participants i.e., PhD students have been largely male and slightly older in this study.

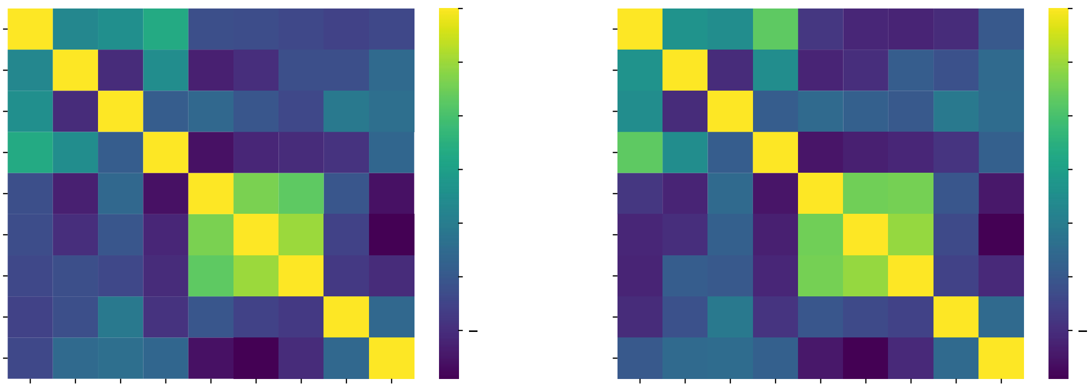

Figure 4-1 A heat map of the correlation coefficients on all of the collected data submitted by the participants. Some parameters like sex or experience had to be normalized likewise to the values of Table 6-1.

Another layer of examination should be the scalability of each method. While this study was conducted with only two parameters in mind, other use cases may require a multitude of parameters. It can be assumed that *1D-Gauss* scales in a linear manner. Meaning, for n parameters this method may require n times a base number of samples. Please note that the necessary sample size of a n-dimensional Gauss fit, already visible in the instability of *2D-Gauss*, would scale exponentially and is thus not practical for higher dimensionalities [19]. Additionally, the scoring of the participants had been less consistent than anticipated during testing. This inconsistent scoring might worsen with an increasing number of dimensions of the Gaussian fit since participants might get overwhelmed by the choices available. *Staircase* could scale significantly worse since multiple passes over all parameters might be required, because of possible influences in between them. A change in one parameter might require an adjustment in a previously set one. *Reference* might confuse inexperienced users especially if the number of possible options is too vast. This is possibly indicated by the slightly stronger correlation between experienced users and *Reference*. Only the *Genetic* method has the possibility to perform better than a linear relationship. However, the population and generation size were set rather arbitrarily, the only goal being a robust and consistent outcome. Additional experiments with genetic algorithms used in a psychometric as opposed to the more frequent machine learning environment, could clarify some rules of thumb when working with said algorithms. Machine learning algorithms often work with hundreds of generations and large starting populations, which is not feasible for human interaction [20]. It might also be of interest to study different genetic algorithms, which have been specifically designed with multi-objective optimization in mind. Of note is the NSGA-II algorithm and its derivatives [21].

4.1 Limitations

No study is without its limitations. For example, the participant pool is mainly consistent of young people, with the mean age being 27. This could affect the average natural ability, individual scene preferences as well as the rating given for every single method. The correlation coefficients (visible in Figure 4-1) are indicative of this behaviour, although this might be skewed by the fact that for this participant pool age highly correlates to experience $\left(r_{\mathrm{p}} = 0.46, r_{\mathrm{s}} = 0.66\right)$.

Another relevant factor might be the loading time of each image. Especially the *Staircase* method has suffered from the fact, that each new image was calculated in real time, which in practice took somewhere between two to three seconds. Participants complained that they had found it difficult to remember the past images. One might negate this effect by either providing more raw processing power or by pre-calculating several images to shorten the time in between images. Although, this pre-calculation of all possible solutions might not be feasible, especially if the number of parameters increases. Additionally, for this method the chroma modifier was multiplied with each recent image, thus raising the absolute step-size for each subsequent image for the increasing direction and vice versa. The change in step size might have also influenced the outcome, bit in practise led almost never to detectable changes. However, this could be easily avoided by calculating a fixed step size with the reference picture at the beginning.

Furthermore, all participants were unaware of the parameters that were being modified except for testing the control method. In practice, most products and/or use cases would give information on these parameters to make it easier for the user. This might benefit some methods more than others. On top of that, research has suggested that there are significant differences in the perception of colour for separate cultural backgrounds [22]. It might also be of interest if those differences extend to the methodology of optimizing for those colours.

Finally, the effects of conducting this test only in a digital form are unknown. Although the experiment was designed to reduce any influences between the room and monitor, a real-world study should be conducted as well. A study in which real objects would be displayed under different lighting conditions might play out differently than just showing images digitally.

5 Conclusion

In conclusion, special care needs to be taken when selecting optimization algorithms. Some methods like the popular *Staircase* might not work well with multiple parameters, because of time constraints and adaptation. Other methods like a Gaussian fit only work effectively for a small number of parameters, since the required sample size scales up to quickly. Genetic algorithms seem to offer the greatest balance of speed and precision i.e., results which don't differ significantly from the ideal. Furthermore, *Genetic* was generally well received for its intuitiveness. Intelligent algorithms, which adapt actively based on already acquired information, are the most promising to optimize multiple parameters. As mentioned before, it might also be of interest to study different and more complex genetic algorithms as well as the required population size and number of generations. Furthermore, since this experiment was only performed digitally its findings might not convert directly into real world examples. A future study with objects under different lighting conditions could confirm the results.

6 Appendix

Table 6-1 All the information gathered on the participants as well as their personal score on how 'intuitive' the tested method was perceived. Experience refers to technical experience in the field. Please note that for the calculation of the mean and standard deviation, 'yes' as well as 'm' have been converted to a 1, while 'no' and 'f' have been converted to 0.

	General Information				
	Age	Sex	Glasses	Colourblind	Experience
Mean	27.25	0.79	0.46	0.00	0.61
Std.	7.65	0.42	0.51	0.00	0.50
	User-Scores				
	1D-Gauss	2D-Gauss	Genetic	Staircase	Reference
Mean	7.29	7.39	7.36	5.46	6.79
Std.	1.12	1.20	1.42	2.03	1.69

Table 6-2 The recorded p-value for each statistical test that was conducted. The significance level was set at α = 5% = 0.05.

	Shapiro-Wilk		Wilcoxon	
	Chroma	Hue	Chroma	Hue
1D-Gauss	0.000	0.001	0.509	0.056
2D-Gauss	0.000	0.000	0.585	0.412
Genetic	0.847	0.010	0.003	0.139
Staircase	0.048	0.699	0.001	0.002
Reference	0.370	0.457	-	-

Table 6-3 Measured colour temperature, luminance and D_{uv} of several surfaces in the test room. All measurements were done using a CS-2000 Spectroradiometer.

	Target	Desk	Wall	Monitor
CCT [K]	6500	6263	6548	6520
Lum. [cd/m^2]	120	110.1	120.7	121.8
D_{uv}	0.000	0.000	0.000	0.002

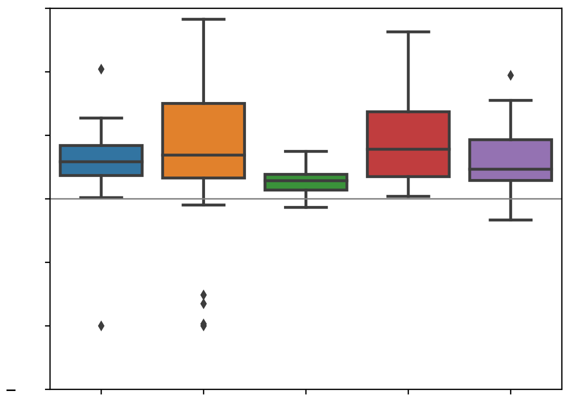

Figure 6-1 Absolute values of the chroma modifiers.

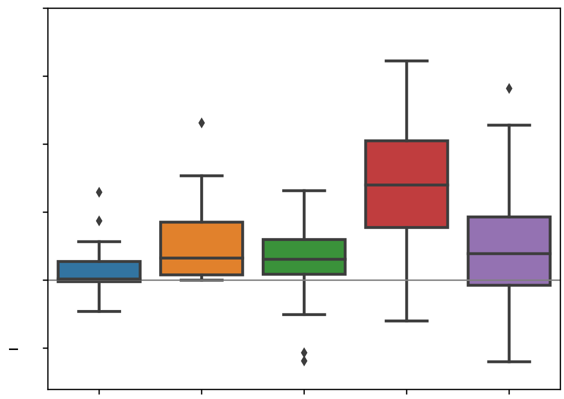

Figure 6-2 Absolute values of the hue modifiers.

7 References

[1] Mark D. Fairchild, *Color Appearance Models: CIECAM02 and Beyond*, 9.11.2004.

[2] K. A. G. Smet, “Tutorial: The LuxPy Python Toolbox for Lighting and Color Science,” *LEUKOS*, vol. 16, no. 3, pp. 179–201, 2020.

[3] Shenzhen Myledy Co Ltd, “Why R9 matter for LED lights?,” 6/28/2017.

[4] Prof. Dr. Stefan Weinzierl, “Labor Kommunikationstechnik: Perzeptive Messung und Evaluation,” Technische Universität Berlin.

[5] M. R. Leek, “Adaptive procedures in psychophysical research,” *Perception & psychophysics*, vol. 63, no. 8, pp. 1279–1292, 2001.

[6] H. Levitt, “Transformed Up-Down Methods in Psychoacoustics,” *The Journal of the Acoustical Society of America*, vol. 49, 2B, pp. 467–477, 1971.

[7] T. Q. Khanh, P. Bodrogi, Q. T. Vinh et al., “Colour preference, naturalness, vividness and colour quality metrics, Part 1: Experiments in a room,” *Lighting Research & Technology*, vol. 49, no. 6, pp. 697–713, 2017.

[8] G. R. Newsham, R. G. Marchand, and J. A. Veitch, “Preferred Surface Luminances in Offices, by Evolution,” *Journal of the Illuminating Engineering Society*, vol. 33, no. 1, pp. 14–29, 2004.

[9] B. R. Shelton and I. Scarrow, “Two-alternative versus three-alternative procedures for threshold estimation,” *Perception & psychophysics*, vol. 35, no. 4, pp. 385–392, 1984.

[10] N. A. Macmillan, “Covert converging operations for multidimensional psychophysics,” *Behavioral and Brain Sciences*, vol. 15, no. 3, pp. 573–574, 1992.

[11] A. Pentland, “Maximum likelihood estimation: the best PEST,” *Perception & psychophysics*, vol. 28, no. 4, pp. 377–379, 1980.

[12] S. Babilon, J. Klabes, P. Myland et al., “Memory colors and the assessment of color quality in lighting applications,” *Optics express*, vol. 29, no. 18, pp. 28968–28993, 2021.

[13] P. Virtanen, R. Gommers, T. E. Oliphant et al., “SciPy 1.0: fundamental algorithms for scientific computing in Python,” *Nature methods*, vol. 17, no. 3, pp. 261–272, 2020.

[14] M. A. Branch, T. F. Coleman, and Y. Li, “A Subspace, Interior, and Conjugate Gradient Method for Large-Scale Bound-Constrained Minimization Problems,” *SIAM Journal on Scientific Computing*, vol. 21, no. 1, pp. 1–23, 1999.

[15] I. Ashdown, “Non-Imaging Optics Design Using Genetic Algorithms,” *Journal of the Illuminating Engineering Society*, vol. 23, no. 1, pp. 12–21, 1994.

[16] J. Blank and K. Deb, “Pymoo: Multi-Objective Optimization in Python,” *IEEE Access*, vol. 8, pp. 89497–89509, 2020.

[17] G. W. Corder and D. I. Foreman, *Nonparametric statistics*: *A step-by-step approach*, Wiley, Hoboken, New Jersey, 2014.

[18] A. Field, J. Miles, and Z. Field, *Discovering statistics using R*, Sage, Los Angeles, CA, USA, 2014.

[19] Julie Delon, *The curse of dimensionality*, 2017.

[20] S. Gotshall and B. Rylander, *Optimal population size and the genetic algorithm*, 2002.

[21] K. Deb, A. Pratap, S. Agarwal et al., “A fast and elitist multiobjective genetic algorithm: NSGA-II,” *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[22] S. Babilon and T. Q. Khanh, “Impact of the adapted white point and the cultural background on memory color assessments,” *Color Research & Application*, vol. 45, no. 5, pp. 803–824, 2020.