https://doi.org/10.2352/issn.2169-2629.2019.27.57

Consistent colour appearance - A novel measurement approach

Marco Mattuschka, Andreas Kraushaar, Philipp Tröster, Jaqueline Wittmann

Abstract

If an image, for example a company logo, is shown on different devices the degree of colour consistency amongst this set of stimuli can be defined as common or consistent colour appearance (CCA). In this work, a procedure which is able to evaluate CCA is developed for the first time. A psychophysical experiment is presented with which the existence of CCA is proofed. For the evaluation of CCA, the colour naming approach from [5] is consistently continued and a measuring tool is developed. In addition, the correctness of the measuring tool is tested on the basis of the experiment. This work is based on two psychophysical experiments, the first to proof CCA, the second to create the colour naming database. This setup is very general and can therefore also be applied to other cultures in order to develop a measuring tool for CCA.

Introduction

Conventional colour rendering methods, which are widely used throughout the graphic arts industry, are generally based on the reproduction of the three integrated parameters X, Y and Z defined by the International Commission for Illumination (CIE). The traditional colour metrics are based on these parameters and the associates CIELab values (ISO 11664-4) and form the basis for most of today's print reproduction processes. Today, there are a multitude of possible production processes with widely varying colour spaces. A typical example is a trade fair stand where a company logo is to be printed on offset, digital, flexographic printing, metal decorating, on a monitor and additionally in 3D printing on widely different media.

Figure 1: A typical example is the trade fair stand, reproduction of a motive in different output types.

As a service provider in the graphic arts industry one aims to harvest the maximal potential of all these reproduction processes with different colour gamuts. Due to this requirement it is insufficient to match a minimum colour difference between the different reproductions on the gamut intersections of the media. Since conventional colour metrics cannot be applied, one-to-one due to the large colour differences to overcome, a somewhat loose relationship to the reference is required if the full potential of each area is to be exploited. Nevertheless, one does not simply want to exhaust the full colour space, but to achieve a "common appearance" (CCA) of the different instances of the artwork. With respect to the seven types of reproduction defined by Hunt [6], which are very much concerned by a pair of stimuli (types 1-6) or by a single stimulus (type 7), the key about common or consistent appearance is that instances of a stimulus from across different conditions need to have something in common. Hence one can think of consistent appearance as being an extension of types 6 or 7, where comparing two stimuli would be the simple case. It should be noted that consistent colour appearance as defined and used here covers the simultaneous evaluation (also known as side by side) of the pertinent samples, compared to a sequential assessment in isolation (also known as media relative). Figure 2 shows the concept of CCA by presenting two sets of images "A" and "B", where it is obvious for many observers that set A has less in common than set "B".

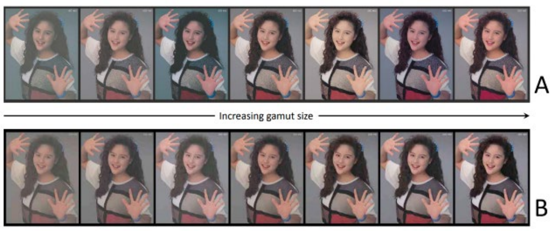

Figure 2: two different reproduction strategies on different output types

Overview of this work

In this work, at first it is determined whether is a such a thing as consistent colour appearance (CCA) and if so, to develop a method to evaluate it. In order to achieve the former, a psychophysical experiment was conducted in which different reproductive strategies had to be evaluated in side-by-side comparison. For the evaluation, the colour name approach from [5] is improved and compared with the results of this experiment.

Experimental setup

Based on the psychophysical experiment from [5], the following experiment focuses on the detection of common appearance, defined as the degree of visual consistency among a set of stimuli. This means: Does any observer always prefer a particular set among several sets of stimuli, because it seems more pleasing or showing more consistency to each other. The simple concept of the experiment is explained in Fig. 3.

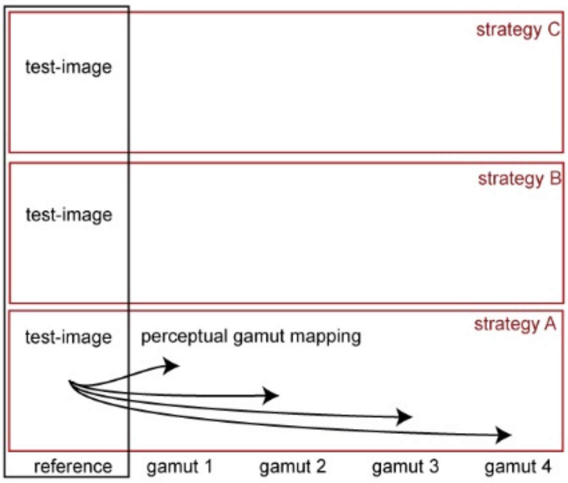

Figure 3: Concept of the experimental setup. A test-image, originating from a reference gamut is mapped to several gamuts via different perceptual mapping strategies.

If someone wants to present an image such as a company logo on several mediums in a way, that they "match each other" usually the perceptual gamut-mapping approach is chosen to map the data. In contrast to the ICC absolute and relative colorimetric rendering intents (RI) this mapping approach is not mathematically defined (ICC SPEC V4). Thus the result depends on the profile maker and is hence "50% art and 50% science" [14]. Therefore, each software package, which offers this RI maps the colours in different ways and claims that their approach results in common appearance. Software solutions from manufacturers which are common in the printing industry are used to investigate uniform colour reproduction. The gamut-mapping strategies used in this experiment are listed in Table 1.

Agfa ColorTune
ColorLogic CoPra
GMG Color
Heidelberg ColorTool
Konica Minolta
X-Rite i1Profiler
X-Rite i1Profiler using black point compensation

Table 1: Software packages performing the perceptual gamut-mapping.

In the experiment, the following typical printing conditions with very different colour gamut were selected: Fogra 53, Fogra 52 (with modified white point), Fogra 39, JapanColor2003WebCoated, CGATS21-2-CRPC1 and PSR-SC.

The previous experiment was based on a test form containing only technical tones, hence colour patches. In order to be able to make a statement about a consistent colour appearance of practice-relevant picture motives, a picture-like test motive is chosen for this project. In order to make it easier for the test persons to compare the resulting reproductions, particular care was taken in selecting the appropriate pictures. In cooperation with the CIE 8-16 working group [13], a collection of test samples was created. With the help of a MATLAB routine developed for this purpose, the images for which the applied gamut mapping strategies deliver the highest possible colour differences (CIEDE2000) were selected from the collection. This should ensure that the visual differentiation of the different reproductions is as easy as possible for the test persons. The following example images were selected by this optimization:

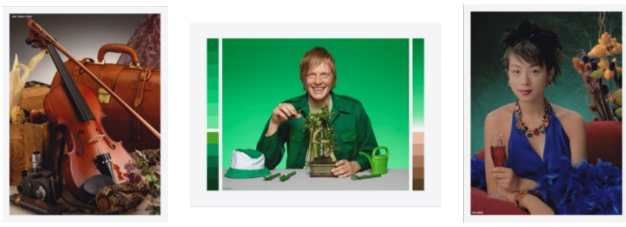

Figure 4: Selected test images.

In total, six reproductions were compared by seven gamut mapping algorithms, i.e. CCA-strategies. This resulted in 21 randomly arranged pair comparisons per experiment. The test picture was determined by the experiment supervisor at the beginning of the experiment and retained during the experiment. A hard proof and a soft proof version of this experiment were performed. For the hardproof experiment, the printed test forms were attached to a neutral grey cardboard according to the arrangement shown and examined by the test person in a standard light booth. In the softproof version, the reproductions were carried out on a calibrated monitor suitable for a softproof [8]. The exact wording can be found here [11].

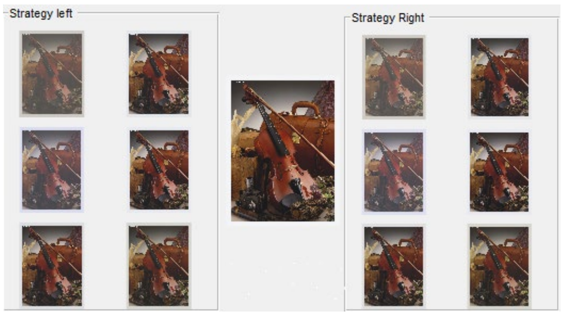

Figure 5: Dialog box for monitor (softproof) experiment. The test person must decide whether the left or the right reproduction is more consistent.

During the experiment, the original image in the middle and a randomly selected set of reproductions of a gamut mapping strategy on the left and on the right side are displayed. The test persons have the task to evaluate which of these sets are more "consistent". In addition to this decision ("left" or "right"), there is also the possibility of a "tie" in the event that the currently displayed samples are not clearly distinguishable in terms of their consistency.

Results of the experiment

A total of 120 people took part in the experiment. Almost every participant has many years of experience in colour matching. The participants were between 16 and 65 years old and all age groups were about the same size. Almost all test persons came from Western Europe. 21 pair comparisons of a test person were evaluated with the following simple procedure: A strategy receives one point for a "won" comparison, a point is subtracted for a lost comparison. A tie is scored with zero points. Thus a strategy can receive a maximum of 6 points. Under the assumption that a strategy wins, loses or is a tie with a probability of 1/3, the expected value for this system is 0 points. These results were averaged over all test persons and are listed in figure 6.

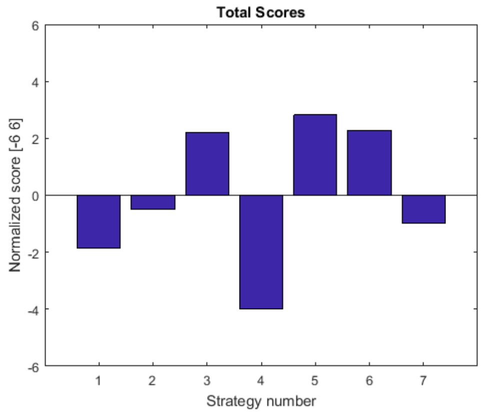

Figure 6: Evaluation of the experiment.

From these data one can conclude that strategies 3,5 and 6 are superior and strategy 4 is clearly inferior. Due to the experimental setup and the number of samples (n=120), significance can be shown for this statement.

Superior	Neutral	Inferior	p-value
3,5,6	1,2,7	4	0.0487

Table 1: Interpretation of the results from Figure 6. The p-value indicates the probability of achieving a similar or more extreme result in a uniformly distributed system.

In this way, this data ensures the existence of CCA. A detailed evaluation that also deals with outliers can be found in [12].

Connecting CCA to colour naming

As briefly stated in the introduction the conventional colorimetry is not able to deal with colour differences as large as present in this study in a meaningful way. However, the field of colour naming seems to be more promising. Therefore, colorimetry does not seem to be a feasible tool to provide a metric for CCA. However, the field of colour naming seems to be more promising. Colour naming is closely related to linguistic relativity, which was made popular in a study of Berlin and Kay [7]. The study deals with the connection of language and thought. There is a debate going on if colour-terminology has absolute universal constraints or if it is dependent on cultural background and language. In order to examine this issue, several monitor-based online-experiments were carried out trying to find out the names of colours, their frequency and their exact position in the colour space.

Two of these studies are the colour naming experiment of Giordano Beretta and Nathan Moroney [9, 10] and the colour naming experiment of Dimitris Mylonas [15]. In both monitor-based experiments, the participants were shown several colour patches and asked to name them. The used sRGB values of the patches are statistically evaluated and associated with a colour-name. The study of Moroney results in 746 names and the Study of Mylonas in 489.

The sRGB triplets can be converted to CIELab values and are assumed to be the centres of volumes in the colour space, which belong to a certain colour name. Basically one ends up with a heterogeneous colour-name-density in the Lab-space. After further transformation into a visually equal colour space, the Din99o colour space [2], a Voronoi diagram was created based on these colour centres. For each value in the colour space, there exists a closest colour centre (colour name). Every value is assigned to the region with the closest centre point. In this way one receives disjoint, convex Voronoi regions V. After limitation to the AdobeRGB colour space these are additionally limited.

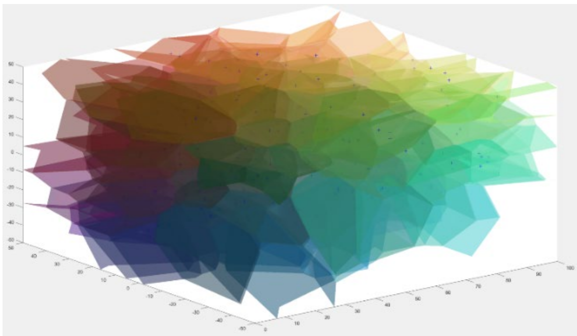

Figure 7: Voronoi colour centres in the Din99o colour space.

A function $d: Lab \times Lab \rightarrow [0, \infty)$ should be developed, which is able to evaluate CCA with the help of a colour name database. The idea is to connect the two input vectors (CIELAB values) by a straight line and to count all intersected colour names.

The trivial solution is to link the two input vectors with a straight line and to count the crossed Voronoi regions V of the diagram. The disadvantage of this solution is, that a colour region can no longer be separated by the metric. Further the function d would not be continuous. Therefore, the following procedure was developed. The Voronoi regions are convex, not empty, limited and connected subsets of $\mathbb{R}^3$. Thus for each Voronoi region V there is a Lipschitz map f, which maps V to the unit ball. Let V be parameterized by the following:

$$V = \{ F(r,\theta,\varphi) = r * (\cos\theta \sin\varphi , \sin\theta \sin\varphi , \cos\theta) + v_m \mid 0 \le r \le r_{max}(\theta,\varphi), 0 \le \theta \le \pi, 0 \le \varphi \le 2\pi \}$$

Here v_m is the Voronoi centre and thus equal to the value of the colour centre, which produces the region. The function $F(r,\theta,\varphi)$ are the spherical coordinates. The mapping f is given by the following rule:

$$f: B_1(0) \to V;$$

$$f(r,\theta,\varphi) = r_{max}(\theta,\varphi) \cdot F(r,\theta,\varphi) + v_m$$

In the following the map f is referred as the trivialisation of the Voronoi region. We notice that this function can be classically differentiated almost everywhere according to Rademacher's theorem.

Let $\gamma: [0,1] \to \mathbb{R}^3$ be a differentiable curve connecting two Din99oLab values. Due to the convexity of the regions, there exists a decomposition $0 \le a_0 < \ldots < a_i < \ldots < a_n = 1$, so that $\gamma([a_{i-1}, a_i]) = \gamma([0,1]) \cap V_i$ for an arbitary enumeration of the Voronoi regions. Thus one can count the overstepped colour names with the following definition:

$$L(\gamma) = \sum_{i=0}^{n} \frac{1}{2} \cdot \int_{a_i}^{a_{i+1}} |Df_i(\gamma)\gamma'|_{ekl} * \theta(f_i(\gamma(t)))dt$$

$$\theta(x) = \begin{cases} 1 \; if \; |x| \le 1 \\ 0 \; else \end{cases}$$

Here f_i is a trivialization of the Voronoi region V_i and $\frac{1}{2}$ is a normalization constant which was chosen as a function of density, so that one is counted if a region is completely exceeded through its centre. With the help of this normalization, the value of this figure reflects the exceeded colour names. Since the definition of L is invariant to the reparameterization of the curve gamma, in this approach the following $d(Lab_1, Lab_2) = L(\gamma) \; with \; \gamma(t) = Lab_1 + t \cdot (Lab_2 - Lab_1)$ is used.

The following example shows how many colour names are pierced by this reproduction of different stimuli of an object.

Lab1	Lab2	CCA-Distance	Overstepped Colournames	Overstepped Colours
55.0 36.0 18.0	48.0 18.0 17.0	1.3	peachy_pink	
			old_pink	
			pale_red	
			pinkish_brown	

Figure 8: The overstepped colour names and the distance for the first density function for poor colour reproduction on a food-packaging.

The density function $\boldsymbol{\theta_i} = \boldsymbol{\theta \; o \; f_i}$ should ideally be the density function of a statistical variable. The density for a Lab colour value expresses the percentage of how many individuals would assign this Lab value to the colour name of the associated Voronoi region. Since the raw data for this density was not available, a distribution was estimated[1]. It was assumed, that every participant would associate a given value with its Voronoi region with probability one. As a result, the density is one. A Voronoi region therefore corresponds exactly to the quantity for which the density is not equal to zero, i.e. $\boldsymbol{x} \in \boldsymbol{V_i} \Leftrightarrow \boldsymbol{\theta_i} = \boldsymbol{1}$.

Another approach is to select this density in the region as one and then let it drop linearly with increasing distance to the centre. As an example the following density was used

$$\theta^2(x) = \begin{cases} 1 \; if \; |x| \le 0.5 \\ 1.5 - |x| \; if \; 0.5 \le |x| \le 1.5 \\ 0 \; if \; |x| > 1.5 \end{cases}$$

Other distributions of this event are also possible here. For both densities an evaluation has been made in the next chapter. For the second density it was additionally assumed that the density of the colour centre belonging to the reference value was weighted with 0.1. This results in a better correlation. However, the function d thus loses its symmetry property and is not a metric by definition. A metric is a function such as d where symmetry and triangular inequality also apply. It should also be noted that the validity of the triangular inequality is also generally not fulfilled. A reasonable definition for the first density could look like this:

$$d(Lab_1, Lab_2) = \min_{\substack{\gamma \in c^1, \\ \gamma(a)=Lab_1, \\ \gamma(b)=Lab_2}} \sum_{i=0}^{n} \frac{1}{2} \cdot \int_{a_i}^{a_{i+1}} |Df_i(\gamma)\gamma'|_{ekl} * \theta(f_i(\gamma(t)))dt$$

However, an implementation of this metric is much more difficult because differential equations would have to be solved. This implementation, along with the "true" density data, could be a starting point for a new project.

Validation with the CCA Experiment

In order to compare the CCA metrics with the experiment data, a target was used which represents a large coverage of the AdobeRGB colour space. Additionally, dominant colours were extracted from the test images and checked. For the test target, a typical print control strip containing technical tones was used. With the help of a large exchange colour space [1] this was converted into the AdobeRGB colour space and used for the evaluation.

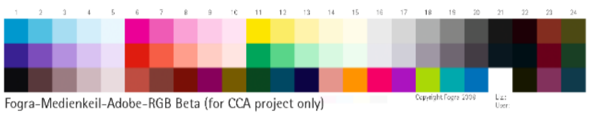

Figure 9: The selected test form for CCA evaluation.

This test target, like the test images, has now been perceptively mapped into the six output conditions using each of the seven strategies. Each individual colour patch of the transformed targets was compared to the reference target using CCA metrics. For each target the 72 distances were averaged on the one hand and on the other hand the maximum was determined. From these values, the average of all six printing conditions per strategy was determined. In order to make the values comparable with the results of

[1] A German based colour naming experiment, led by D. Mylonas, is under way and it is expected to access the individual and the resulting observer data later.

the experiment, the deviation from the mean value was displayed and assigned to the respective strategies in a bar chart. A rather comparable result can be seen when comparing the test pictures of the experiment. For this purpose, the relevant colours of the test images were determined using the k-Means algorithm and the colour centres. The k-Means algorithm requires a given number of clusters. For each test pattern, the k-Means algorithm was started with two clusters. The number of the clusters was inductively increased by one until two cluster centres with the same colour name were assigned.

Figure 10: The extracted relevant colours of the test image.

On the basis of these extracted colours, the evaluation of the seven strategies was carried out equivalent to the CCA test target. The result for this picture evaluation is shown in figure 11.

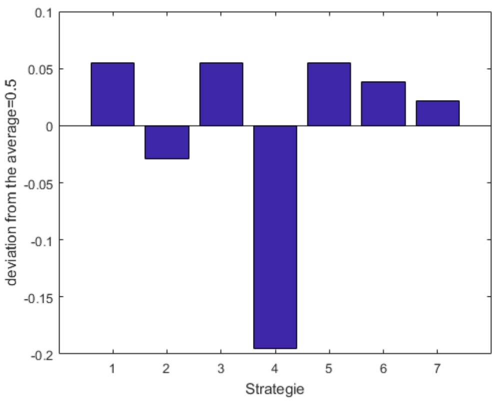

Figure 11: The evaluation of the CCA experiment. Compared to the evaluation of the test picture (see figure 9) a correlation can be seen.

Apart from Strategy 1, there is a correlation between the respondent survey and the CCA metrics. This leads to the conclusion that the approach makes sense. The correlation for the second density function looks still better.

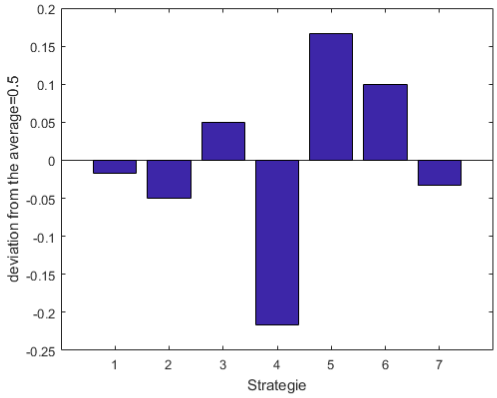

Figure 12: The evaluation of the CCA experiment. Compared to the evaluation of the test picture (see figure 9) a correlation can be seen.

A very comparable statement as shown in table 1 can be made for this result. The evaluation of the CCA test target is very close to the evaluation of the test picture.

Superior	Neutral	Inferior
5,6	1,2,3,7	4

Table 2: Interpretation of the results from figure 11.

Conclusion

Based on the data situation it can be assumed that the presented colour naming approach is a useful method to evaluate CCA. However, this procedure requires a solid database, the colour names database. So a more plausible result can be achieved by a better analysis of the colour names. Another interesting question would be whether this approach would produce a similar result for other cultures. The approach presented here can be tested for other culture groups by performing the experiment in it and using an associated colour name database.

Through the CCA metrics developed in this work, the distance of a reproduction to its reference can be meaningfully evaluated. In this case, the task requires 6 different reproductions to be evaluated simultaneously for consistency to a reference. This was solved in this paper by selecting the worst reproduction of the 72 colour values for each strategy on the basis of the media wedge or colour values extracted from the test images. The mean value of the 6 printing conditions was then calculated from these figures in order to compare the strategies. However, this is only one possibility to evaluate a set of reproductions and is only conditionally meaningful for this task. A meaningful continuation of the work would be to develop at this point a more suitable method for the simultaneous evaluation of different reproductions.

ACKNOWLEDGMENTS

The project "Common Colour Appearance " [Fogra No. 19348] was supported within the programme of the Federal Ministry of Economics and Technology for the support of “Industrielle Gemeinschaftsforschung [IGF]” via the German Federation of Industrial Research Association [AiF/No. 19348 N].

References

[1] A. Kraushaar, New CMYK-based colour exchange space: FOGRA53 and eciCMYK released, URL: https://www.fogra.org/en/fogra-research/wc-digital-printing/digital-printing-current-projects/consolidating-standardization/exchange-colour-space/new-cmyk-based-colour-exchange-space-fogra53-and-ecicmyk-released.html

[2] Norm DIN 6176: 2001, Farbmetrische Bestimmung von Farbabständen bei Körperfarben nach der DIN99-Formel, How to get: Beuth-Verlag, 10772 Berlin & G. Cui; M. R. Luo; B. Rigg; G. Roesler; K. Witt Uniform Colour Spaces Based on the DIN99 Colour-Difference Formula; In: Color Research & Application (2002), Volume. 27, Nr. 4, S. 282-290

[3] Kaufmann R., O'Neill M. C..: *Color names and focal colors on electronic displays.* Ergonomics, pp. 881-890, (1993).

[4] McFadden S., Kaufmann R., Janzen M.: *Classification of colors presented against different colored backgrounds.* Displays, **15**, pp. 203-214, (1994).

[5] Tröster, P; ET. AL.: A pilot study on evaluating common appearance and a colour naming approach to measure it, CIC 2015, Darmstadt

[6] Hunt RWG. The Reproduction of Colour, 6th edition. Chichester & Fairchild M.: Colour Appearance Models. Wiley & Sons, West Sussex - England, pp. 310 ff.

[7] Berlin B., Kay P.: *Basic Color Terms: Their Universality and Evolution*. CSLI Publications, Stanford, California, (1999)

[8] Einflussgrößen auf die hochqualitative Monitordarstellung hinsichtlich einer Standardisierung von Softproofsystemen. München: Fogra (2014) (Nr. 10.058) – Forschungsantrag

[9] Beretta G. B., Moroney B. M.; Hewlett-Packard Laboratories, pp. 1-27, (2010).

[10] Moroney N.; Electronic Imaging pp. 36-46, (2003).

[11] Project webpage: https://www.fogra.org/fogra-forschung/medienvorstufe/schwerpunkte-medienvorstufe/commonappearance/commonappearance.html

[12] T. Gancheva: Consistent Color Appearance psychometric evaluation using pair comparison method for soft-proof and hardproof paper simulated samples under different strategy https://www.fogra.org/fogra-forschung/medienvorstufe/schwerpunkte-medienvorstufe/commonappearance/publikationen-ca/publikationen.html

[13] 8-16: Consistency of Colour Appearance within a Single Reproduction Medium http://www.cie.co.at/technicalcommittees/consistency-colour-appearance-within-single-reproduction-medium

[14] Morovič J.: *Color gamut mapping*. John Wiley & Sons. USA, (2008).

[15] Mylonas, D., & MacDonald, L.; Color in Graphics, Imaging. Joensuu, Finland: IS&T (2010).

Author Biography

Marco Mattuschka received his diploma in mathematics (2012) and his PhD in mathematics from the Albert Ludwig University of Freiburg (2018). Then he joined the Fogra research institute in Munich where he works as scientist and developer. Amongst other topics as 3D-printing and artificial intelligence, he is doing research in the field of applied colour-science.