Validation of Quality Attributes for Evaluation of Color Prints

Marius Pedersen*, Nicolas Bonnier**, Jon Y. Hardeberg* and Fritz Albregtsen.***
**** Gjøvik University College (Norway). ** Océ Print Logic Technologies S.A. (France). *** University of Oslo (Norway).***

Abstract
Image quality assessment is a difficult and complex task. Quality attributes have been used in the evaluation of perceived image quality in an attempt to reduce the complexity and the dimensionality. Recently, Pedersen et al. (CIC, 2009) proposed a set of quality attributes for the evaluation of color prints. In this paper we perform an experimental validation of these quality attributes to ensure that the criteria on which they were selected are fulfilled. The results show a correspondence between the quality attributes and the criteria. The quality attributes are therefore considered as a good starting point to describe overall image quality.

Introduction
New and more advanced products are introduced continuously into the printer market. One of the reasons for these rapid technology advancements is the wish to produce high-quality prints fast and economically. In order to verify that the technology advancements produce prints of higher quality than with the current technology, some kind of quality assessment is required.

There are two main classes of methods to assess Image Quality (IQ), subjective and objective. The subjective is carried out by human observers, while the objective does not involve observers. Objective assessment can involve the use of measurement devices to obtain numerical values, alternatively IQ metrics can be used. These IQ metrics are usually developed to take into account the human visual system, and thus with the goal of being correlated with subjective assessment. Our long term goal is to create a link between subjective assessment and objective IQ metrics.

Both subjective and objective IQ assessment are dependent on a number of Quality Attributes (QAs), which are terms of perception, such as sharpness, lightness, and saturation [1]. IQ is influenced by these QAs, and knowledge about the different QAs and how they influence IQ can be used to achieve high-quality reproductions and to help in the IQ assessment.

In earlier papers [2, 3] we proposed a set of six QAs for the evaluation of color prints: color, lightness, artifacts, contrast, sharpness, and physical. We refer to these as the Color Printing Quality Attributes (CPQAs). These CPQAs originated from a literature survey of the existing QAs used by researchers to evaluate IQ. They were selected based on different requirements:

- the QAs should be based on perception,
- they should account for most technological issues,
- they should be straightforward to use,
- they should be suitable for IQ metrics,
- they should be as independent as possible,
- the number of QAs should be kept to a minimum (low dimensionality),
- they should be useful for the evaluation of color prints.

In a preliminary experiment almost all the QAs used by a group of observers were grouped within the proposed CPQAs [2, 3]. This preliminary validation was done by subjectively grouping the numerous QAs used by the observers to one of the six CPQAs. However, since it was carried out subjectively by the authors, it does not fully validate the aspects on which the CPQAs were selected. Additional validation is therefore required to ensure that they satisfy the intended needs, and to make sure that the proposed CPQAs are suitable to assess IQ. The goal of this paper is to validate the CPQAs experimentally.

This paper is organized as follows: First a section on how to validate QAs. Next the experimental setup, followed by a section on grouping the data from the experiment to the CPQAs. Further, a part on observations on the CPQAs made during the experiment. At last, we summarize the validation, conclude, and propose future work.

How To Validate Quality Attributes?
The validation should be adapted to the criteria on which the CPQAs were selected. The validation can be achieved by comparing data to the CPQAs, and analyzing the correspondence between the data and the CPQAs. Requirements need to be set to validate the CPQAs. Using the aspects on which they were selected (the above bullet point list) we can derive the important requirements that the CPQAs should fulfill. For the CPQAs to be useful for the evaluation of IQ, to be perceptual, and account for technological issues, they should be able to cover the entire field of IQ. All issues encountered in the evaluation of color prints should be described using the CPQAs, making this one of the requirements to validate. As many as possible of the QAs used by the observers should be accounted for within one of the CPQAs, and not overlap several CPQAs. Minimum overlapping is considered as one of the requirements the CPQAs should fulfill. The CPQAs were selected to keep the number of QAs to a minimum, this is important for usability of the QAs, and for the CPQAs to be straightforward to use. Therefore dimensionality should be one of the requirements. For the CPQAs to be suitable for IQ metrics and straightforward to use, it is important to keep independence.

Summarized, we have four different requirements the CPQAs should fulfill in order to be validated:

- the CPQAs should cover the entire field of IQ,
- few QAs should overlap the CPQAs (i.e. most of the QAs can be assigned to only one of the proposed CPQAs),
- dimensionality should be kept to a minimum,
- low or no dependence should occur between CPQAs.

There are several ways to carry out the validation for these requirements. The validation can be carried out subjectively or objectively. The drawback of the previous validation [2, 3] of the CPQAs was subjectivity. In order to minimize the subjective influence, and to have an independent validation of the QAs; objective validation methods have been investigated. It is preferable to have a fully objective method, where data, for example from an experiment, can be compared to the CPQAs. This requires a database containing all QAs, categorization of them, and their relations. To our knowledge such a database does not exist, and this method is therefore inapplicable. Another possible method is to use existing definitions of QAs to create relations between the QAs, resulting in a data structure, which can be visualized as a tree or data structure. This method is not completely objective, but it keeps the subjectivity to a minimum. Therefore, we will adopt this method for the validation of the CPQAs.

Since the CPQAs are based on human visual perception, subjective data is required for the validation. In order to validate if the CPQAs cover the entire field of IQ it is required that the observers use a wide variety QAs. Expert observers have been shown to be more precise than non-experts [4] and they have a wider vocabulary. Therefore expert observers should be recruited for such experiments. In addition, the color workflow on which the data is collected should guarantee many different quality issues. The image set should also include a wide variety of characteristics to ensure that many different IQ issues are encountered.

There are various ways to carry out such an experiment. One way is to provide the CPQAs and their definitions to the observers, and ask the observers to use them in their judgment of IQ. If the observers only use the CPQAs, one could argue that they cover all aspects of IQ. However, this experimental setup can restrict the observers to the CPQAs, and prevent them from using others QAs they normally would use. Another option is to record the QAs used by the observers during the experiment, where the observers do not have prior knowledge to the CPQAs. This last option does not restrict the observers to the CPQAs. Therefore, it has the best potential to validate the CPQAs, we adopt this method.

Experimental setup

Images

Several guidelines have been given in the literature for the selection of images, in the context of investigating IQ issues. Holm et al. [5] recommend the use of a broad range of natural images as well as test charts to reveal the quality issues. The Commission Internationale de l'Éclairage [6] suggests to include images with the following characteristics: high-key, low-key, low lightness contrast, leaves and sky, no neutrals, no white point, no black point, heavy cast, few hues, business graphic, and flesh tones. Büring et al. [7] propose to use natural images, as well as saturated colors. In this experiment we have selected 25 images (Figure 1), which are chosen based on different image characteristics:

- low, medium, and high levels of lightness,
- low, medium, and high levels of saturation,
- hue primaries,
- low, medium, and high contrast,
- larger areas of the same color,
- fine details,
- memory colors such as skin tones, grass, and sky blue,
- color transitions,
- neutral gray.

To address the customer segment of Océ, we have also included 3D models, maps, posters, presentations, and pdf-like documents. The images have been collected from different sources. One image from ISO [8], two from CIE [6], ten images from the authors, one image from MapTube [9], three images from ESA [10], four images from Google 3D Warehouse [11], one image reproduced with permission from Ole Jakob Skattum, and one image from Halonen et al. [12]. The images were 150 dpi 16-bit sRGB, saved as tiff files without compression.

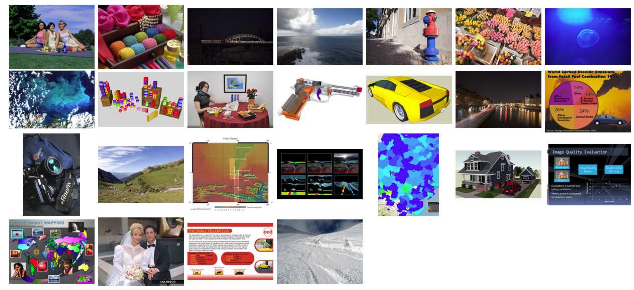

Figure 1. The 25 images used in the experiment to validate the quality attributes.

Color workflow

The images were printed on an Océ Colorwave 600 CMYK wide format printer on Oce Red Label (LFM054) plain uncoated paper. The profile of the printer was generated using a GretagMacbeth TC3.5 CMYK + Calibration test chart in ProfileMaker Pro 5.0.8. A round trip test was carried out to ensure a correct profile as suggested by Sharma [13], and we performed a visual inspection of color gradients to verify that no artifacts occurred. The images were printed with three different rendering intents: perceptual, relative colorimetric, and relative colorimetric with black point compensation.

Viewing Conditions

The observers were presented with a reference image on an EIZO ColorEdge CG221 display at a color temperature of 6500K and a white luminance level of 80 cd/m^2, following the specifications of the sRGB. The image set being rendered for sRGB display, a monitor capable of displaying the sRGB gamut was the most adapted reproduction device. In addition, the display was fitted with a monitor hood to prevent glare. The printed images were presented in random order to the observers in a controlled viewing room at a color temperature of 5200K, an illuminance level of 450 ± 75 lux and a color rendering index of 96. The observers viewed the reference image and the printed image simultaneously from a distance of approximately 60 cm. The experiment followed the CIE guidelines [6] as closely as possible.

Instructions

The instructions given to the observers focused both on the overall quality rating and on the QAs used in the evaluation:
Rank the reproductions according to quality.
- Elaborate on the attributes you use and quality issues you observe, i.e. all attributes you consider.
- If possible try to give an indication of the importance of the issues and attributes, and important areas.

The entire experiment was filmed, and the observers were encouraged to describe and talk about their observations. The video enabled the authors to better capture the attributes used by the observers than if they were to write down the attributes, since observers are usually more articulate orally.

Fitting the QAs Data to the CPQA

Four observers, all considered to be experts, were recruited for the experiment. This resulted in a total of 100 observations by the four observers for the 25 images, and more than six hours of video were recorded. The video was transcribed by the authors with focus on the QAs used by the observers. Numerous QAs, more than 750 in total and more than 350 different QAs, were used by the expert observers. This data constitutes the basis for the validation of the CPQAs. Figure 2 shows a tag cloud of the top 25 words from the raw transcribed data [14].

Figure 2. Tag cloud with the 25 top words from the raw transcribed data. The font size of a tag in the tag cloud is determined by the number of times the word has been used by the observers. Similar words have been grouped, such that details and detail are counted together as detail.

Since many of the words and phrases from the observers are similar and some synonyms, the QAs from the experiment need to be categorized. Similar words and phrases should be grouped together, and relations between terms found. We have chosen to use existing definitions to accomplish this, and two different approaches can be taken with this method; top-down or bottom-up. In the top-down approach the relations are built from the most general QAs and downwards to the most specific QAs. This requires building a full tree structure with all relations, and then comparing it to the QAs used by the observers. In the bottom-up approach, the starting points are the QAs used by the observers. These QAs are grouped into more general attributes till the most general QA is reached. The advantage is that it does not require building a full tree structure prior to the analysis. Therefore, the bottom-up approach was chosen to validate the CPQAs.

An example of how the analysis is carried out; the observer has used the QA hue shift, this QA belongs to the more general QA hue. Using the relations of Pedersen et al. [2, 3] and the definition by Wyszecki and Styles [1], hue is considered a part of the more general color QA, which is one of the CPQAs (Figure 3).

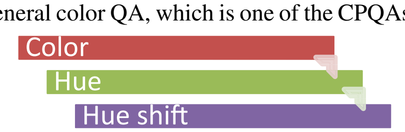

Figure 3. Bottom-up procedure for the attribute hue shift, which belongs to the hue attribute, which in turn belongs to the color attribute.

In the experiment observers described quality issues in the reproductions, differences between the original and the reproductions, and differences between the reproductions. Since the physical CPQA was not changed in the experiment, we limit the discussion to five of the six CPQAs, excluding the physical CPQA.

The bottom-up approach described above has been used to generate a tree for all the images and observers in the experiment (Figure 4). In the following, we will show how the QAs from the expert observers have been grouped and fitted to the CPQAs, but because of page limitations we cannot show the complete process, and will only present a part.

The observers have used many specific terms regarding color, such as red hue shift, yellow hue shift, and blue hue shift. All of these terms indicate a hue shift, which is a child of the hue attribute. A drift in hue also indicated a hue shift. Color dominance and color cast were used to indicate a general hue shift.

The observers specifically indicated which colors had an increase in saturation. They also tended to indicate which image was more saturated than another, rather than the other way around. Also, saturation loss was used in a more general way to indicate a global loss, while saturation increase was often used for a color or a region. Increase and loss of saturation are considered by the authors to be a shift in saturation. Shift in saturation is a sub-QA of saturation. Observers also used the following terms to describe saturation: intensity, chroma, purity, colorfulness, and vividness. Chroma is used for saturation in the Munsell color system [15], in the ISCC-NBS lexicon vividness is a level of saturation [15], purity is also a synonym for saturation [16], colorfulness is considered the same as chroma [16, 17], and intensity is used about saturation density [18]. Based on these definitions these terms are considered as equal to saturation.

Saturation and hue are considered as children of the color attribute. For the general color attribute observers used terms as color shift, color reproduction, and color rendering. In addition, the observers used the term chromaticity. Since the definition of chromaticity by Oleari [19] contains both hue and saturation we can set this attribute as equal to the color attribute.

Discussion on the Fitting of QAs

Several issues were encountered while fitting the QAs.

Overlapping QAs Some of the QAs used by the observers are difficult to group within only one of the CPQAs. Naturalness is one of these attributes. We have argued that naturalness could be accounted for by using several of the main or sub-attributes [2, 3]. In this experiment the observers used several QAs together with naturalness, very often a change in one or several of the other QAs lead to the impression of an unnatural image. In the five observations, where naturalness was used, the observers used the term color in all of them, contrast in three of the five, and memory colors in four of the five observations. In addition, it has been shown that naturalness depends on chroma and colorfulness [20], contrast [20], and memory colors [21]. Because of this, naturalness is most likely accounted for if these QAs are of reasonable quality.

The word gamut was also used by the observers, which is defined as the range of a set of colors [22]. Gamut cannot be listed as a sub-QA under one of the CPQAs, since it is dependent on both the color CPQA and the lightness CPQA. In the three observations where gamut was used, both the lightness and the color QAs were used. Therefore, gamut can be accounted for using the color and lightness CPQAs.

Readability and legibility are two terms from the experiment, which have been found to be related in the literature [23], and they are often used about textual information. Research has shown that contrast is important for text readability [24] and text legibility [25], these terms will also be influenced by sharpness. In five of the eleven observations where legibility and readability were used, the observers used also contrast and sharpness, in the remaining six observations either sharpness or contrast was used. This indicates that legibility and readability most likely can be accounted for with the CPQAs.

Memory colors are placed under the color CPQA, as the observers only specified color changes (saturation and hue) for

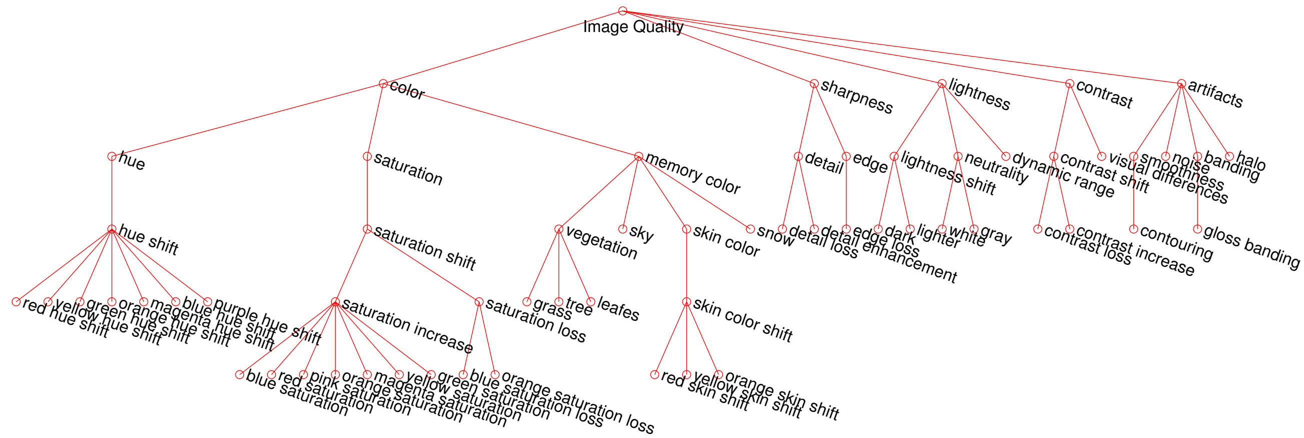

Figure 4. *The QA tree generated from the attributes used by four expert observers. Each level of a sub-tree has been sorted from left to right based on frequency (high to low).*

these, and not changes in lightness. However, there might be situations where lightness should be considered as well, and memory colors in terms of lightness will become a sub-QA of lightness.

Independence Dynamic range is considered as a sub-QA of lightness, but it has also been shown to influence contrast [26]. In the two observations with dynamic range, the observers indicated a relation to visibility of details. This issue is linked to the use of the phrase too dark, which was often used together with detail loss. In these cases, the observers perceived the regions where shadow details were lost, as larger dark uniform areas compared to the original, and used the term too dark or darker.

The experimental data indicates that contrast is influenced by saturation and lightness, but also that contrast is linked to detail. Since the definition of contrast contains both color and lightness it is perhaps the least independent QA. Furthermore, the experimental data shows that the observers often use the contrast attribute separated from the color and lightness attributes, making contrast a very complex QA. Contrast is also important to account for both naturalness and readability. Without the contrast CPQA we would not cover the whole field of quality, and it is therefore required in order to fulfill the criteria on which the CPQAs were selected, even at the expense of independence.

We carried out a cross-tabulation and chi-square analysis to investigate the dependence between the CPQAs. The null hypothesis H_0 was that there was no relationship between two CPQAs. The alternative hypothesis H_1 was that there was a relationship between two CPQAs. The p-values from this analysis are shown in Table 1. For some combinations of two CPQAs given a 5% significance level, H_0 was rejected in favor of H_1. The input data to the analysis was whether or not one of the five CPQAs was used by the observers for each of the 25 images. The disadvantage of this analysis is that it does not give any information on the nature of the relationship between the CPQAs, it only gives information about when two CPQAs are used together. However, from the results we see a dependence between artifacts and lightness, which was also found eariler [2, 3]. There is also dependence between artifacts and sharpness, and contrast and lightness. The observers indicated a relation between contrast and dynamic range, one of the sub-QAs of lightness. The dependence analysis between dynamic range and contrast did not reveal a relation, neither for detail visibility and dynamic range, nor for detail loss and dark. The reason for this might be the amount of data, four observers is too low for these specific QAs (Figure 6). Anyhow, since this analysis does not cover the nature of the relations, further investigation is required to investigate the dependence between the CPQAs, where information on the rating of each CPQA is needed. This information could not be gathered in our experiment since the observers did not have prior knowledge about the CPQAs.

Table 1: P-values from cross-tabulation and chi-square analysis. With a significance level at 5%, there is a dependence between artifacts and lightness, contrast and lightness, and barely between artifacts and sharpness.

	Color	Sharpness	Lightness	Artifacts	Contrast
Color	0	0.405	0.568	0.781	0.423
Sharpness	0.405	0	0.198	0.048	0.230
Lightness	0.566	0.198	0	0.014	0.047
Artifacts	0.781	0.048	0.014	0	0.764
Contrast	0.423	0.230	0.047	0.764	0

Global and Local Issues During the experiment the observers looked at both global and local quality issues. The QAs above can be divided into global and local attributes, this differentiation can be important for the assessment of IQ, but also in the method used to combine results from different QAs.

One Child with Several Own Children In the tree structure (Figure 4) some QAs have only one child, and this child has several own children. It could then be argued that these QAs could be discarded, and the QA below could be linked to the parent of the removed QA. For example, saturation could be removed and replaced with saturation shift. However, observers have used these terms, as saturation alone, without further specification, which indicates that these levels are important and should be kept. Furthermore, in other situations there might be several children, such as for the edge QA, where one could suggest having two children as for the detail QA, one for edge loss and another for edge enhancement.

Skewness The experimental data identifies skewness in terms of the number of sub-QAs between the different CPQAs, which was also found by Pedersen et al. [2, 3]. Our experimental data shows that the color CPQA has significantly more sub-QAs than the other CPQAs. This can be used as an additional argument for separating lightness from the color CPQA, in order to reduce skewness. Additionally, separating these enables the CPQAs to be easily adapted for the evaluation of grayscale images. The disadvantage of skewness between the CPQAs is that it is not straightforward to combine IQ values from the different CPQAs to one overall IQ value, since the CPQAs might have unequal weight.

Dimensionality Since all of the CPQAs have been used by the observers in the experiment, none of the CPQA can be removed directly to reduce the number of CPQAs. However, the color and lightness CPQA could be merged, but at the cost of increased skewness. There were 39 observations where the observers used both color and lightness, indicating that the observers differentiate between these CPQA. There were also nine observations where lightness was addressed without color.

It is not unlikely that for specific workflows with specific documents that the dimensionality can be reduced. Therefore we have also looked at the usage of the CPQAs for the non-photographic documents. For these images all the CPQAs have been used, and for this workflow none of the CPQAs can be removed. However, there might be situations where only a part of the CPQAs are used to evaluate IQ.

Observations on the CPQAs

The experimental data also leads to different observations on the CPQAs, which can be valuable in the assessment of IQ. Figure 5 shows the frequency of use of the CPQAs in the experiment. Color is the CPQA used most frequently by the observers, closely followed by sharpness. Artifacts is the least used CPQA by the experts. The results here indicate that the color and sharpness CPQAs should be evaluated in all images for IQ assessment. The low number of observations regarding artifacts could indicate that this CPQA only needs to be accounted for in specific images, since the artifacts might be dependent on the characteristics of the image. One example is banding, which has been perceived by some observers in the images with large uniform areas, but not in the other images.

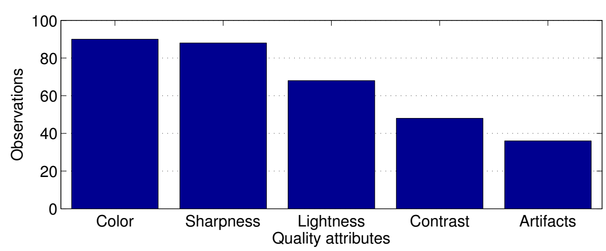

Figure 5. Frequency of use for the CPQAs for the 100 observations. Color is the CPQA used the most by the observers.

It is also interesting to look at the distribution of sub-QAs within the CPQAs. Detail is the most often used sub-QA (Figure 6), closely followed by hue. Since the observers paid much attention to loss of detail in the shadow regions, and since the rendering intents reproduced these regions differently, detail is not surprisingly the most used sub-QA. The perceptual rendering intent gave a slight hue shift in some of the images, which was often noticed by the observers, resulting in the frequent use of this attribute. The sub-QAs of artifacts are the least used, most

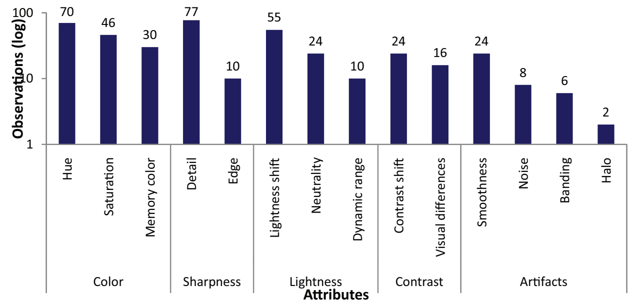

Figure 6. Distribution of sub-QAs within the CPQAs. Detail is most used.

likely since these are very specific. The artifact CPQA will contain many sub-QAs since it will cover many different artifacts.

It has been suggested that the first quality issue noticed by the observer is likely to be the most important. We have analyzed this aspect. Figure 7 shows that color is by far the most frequent first attribute used by the observers, dominating more than in the frequency table for the whole experiment (Figure 5)

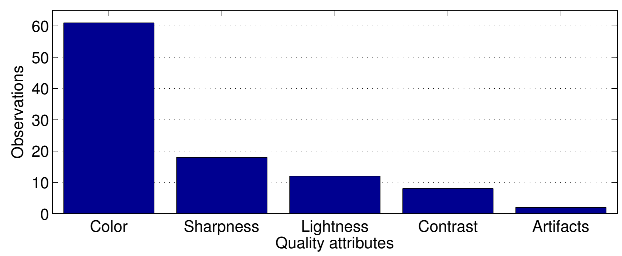

Figure 7. Number of observations based on the first CPQA used by the observers.

Validation Summary

Prior to the experiment, we specified four requirements for the CPQAs to be validated. First the CPQAs should cover the entire field of IQ. This is fulfilled if all the QAs recorded in the experiment can be fitted within one or several of the CPQAs. This requirement is satisfied, and all the recorded QAs are accounted for within the CPQAs, either directly as a CPQA or as a sub-QA, or by using two or more of the CPQAs.

The second requirement was to have as few overlapping QAs as possible. Some of the recorded QAs overlap, such as naturalness, gamut, readability, and legibility. These overlapping QAs have been used totally 15 times, only a small percentage of the total number of QAs used. The overlapping QAs can be accounted for using two or more of the CPQAs. We consider the number of overlapping QAs to be acceptable, and the overlapping QAs are not frequently used by the observers. Thus the CPQAs satisfy the second requirement.

The third requirement was to keep the dimensionality to a minimum. None of the CPQAs can be directly removed, and all CPQAs have been used by the observers. However, as discussed above the division between color and lightness has advantages and disadvantages. They could possibly be combined into a single QA. Nonetheless, without merging color and lightness, and considering the use of only lightness by the observers, the third requirement is satisfied.

The last requirement regarded dependence. The experimental results show some dependence between CPQAs, but as stated by Pedersen et al. [2, 3] the CPQAs are not fully independent, because it is very difficult, if not impossible, to account for all quality issues while maintaining a low dimensionality. The experimental results indicate that contrast is the least independent CPQA. However, contrast cannot be removed since it is often used by the observer. For that reason we consider the dependence found to be acceptable, but care must be taken if the quality values for each CPQA are combined into an overall IQ value.

Conclusion

Five of the six CPQAs proposed by Pedersen et al. [2, 3] for the evaluation of color prints have been validated. An experiment with expert observers was carried out to investigate QAs, where the experimental data acted as the basis for the validation. Four requirements were set for the CPQAs to be considered as appropriate for the evaluation of color prints. The experimental results show that the CPQAs satisfy these requirements.

Further investigation of the dependence between CPQAs is considered as possible future work, together with examination of methods to combine the CPQAs into an overall IQ value.

Acknowledgments

The author hereof has been enabled by Océ-Technologies B.V. to perform research activities which underlies this document. This document has been written in a personal capacity. Océ-Technologies B.V. disclaims any liability for the correctness of the data, considerations and conclusions contained in this document.

References

[1] G. Wyszecki and W.S. Styles. *Color Science, Concepts and Methods, Quantitative Data and Formulae*. Wiley Interscience, Derby, UK, 2nd edition, Aug 2000.

[2] M. Pedersen, N. Bonnier, J. Y. Hardeberg, and F. Albregtsen. Attributes of image quality for color prints. *Journal of Electronic Imaging*, 19(1):011016–1–13, Jan 2010.

[3] M. Pedersen, N. Bonnier, J. Y. Hardeberg, and F. Albregtsen. Attributes of a new image quality model for color prints. In *Color Imaging Conference*, pages 204–209, Albuquerque, NM, USA, Nov 2009. IS&T.

[4] F. Dugay, I. Farup, and J. Y. Hardeberg. Perceptual evaluation of color gamut mapping algorithms. *Color Research & Application*, 33(6):470–476, Dec 2008.

[5] J. Holm, J. Tastl, and T. Johnson. Definition & use of the ISO 12640-3 reference color gamut. In *Color Imaging Conference*, pages 62–68, Scottsdale, AZ, 2006. IS&T/SID.

[6] CIE. Guidelines for the evaluation of gamut mapping algorithms. Technical Report ISBN: 3-901-906-26-6, CIE TC8-03, 156:2004.

[7] H. Büring, P.G. Herzog, and E. Jung. Evaluation of current color management tools: Image quality assessment. In *The European Conference on Colour Graphics, Imaging and Vision*, pages 459–462, Aachen, Germany, Apr 2004.

[8] ISO. ISO 20462-3 photography - psychophysical experimental methods to estimate image quality - part 2: Quality ruler method, jul 2004.

[9] MapTube. http://maptube.org/, 2010.

[10] European Space Agency. www.esa.int, 2010.

[11] Google 3D Warehouse, 2010. `http://sketchup.google.com/3dwarehouse/`.

[12] R. Halonen, M. Nuutinen, R. Asikainen, and P. Oittinen. Development and measurement of the goodness of test images for visual print quality evaluation. In S. P. Farnand and F. Gaykema, editors, *Image Quality and System Performance VII*, volume 7529, pages 752909–1–10, San Jose, CA, USA, Jan 2010.

[13] A. Sharma. Measuring the quality of ICC profiles and color management software. *The Seybold Report*, 4(20):10–16, Jan 2005.

[14] TagCrowd. 07/04/10: http://tagcrowd.com/.

[15] T. Berk, L. Brownston, and A. Kaufman. A new color-naming system for graphics languages. *IEEE Computer Graphics and Applications*, 2:37–44, 1982.

[16] L. Wang, J. Giesen, K. T. McDonnell, P. Zolliker, and K. Mueller. Color design for illustrative visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14: 1739–1754, 2008.

[17] E. J. Giorgianni and T. E. Madden. *Digital Color Management: Encoding Solutions*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2nd edition, 1998.

[18] M. C. Stone. Color in information display principles, perception, and models. International Conference on Computer Graphics and Interactive Techniques archive. ACM SIGGRAPH 2004 Course 20 Notes, 2004.

[19] C. Oleari. *Misurare il colore*. Hoepli, Milano, Italy, Oct 2008.

[20] T.J.W.M. Janssen. *Computational Image Quality*. PhD thesis, Technische Universiteit Eindhoven, 1999.

[21] S.N. Yendrikhovskij. *Color reproduction and the naturalness constraint*. PhD thesis, Technische Universiteit Eindhoven, 1998.

[22] G. Sharma. *Digital Color Imaging Handbook*. CRC Press, Inc., Boca Raton, FL, USA, 2002.

[23] S. Zuffi, P. Scala, C. Brambilla, and G. Beretta. Web-based versus controlled environment psychophysics experiments. In L. C. Cui and Y. Miyake, editors, *Image Quality and System Performance IV*, volume 6494 of *SPIE proceedings*, page 649407, San Jose, CA, Jan 2007.

[24] K. Knoblauch, A. Arditi, and J. Szlyk. Effects of chromatic and luminance contrast on reading. *Journal of the Optical Society of America A*, 8:428–439, Feb 1991.

[25] G. E. Legge, D. H. Parish, A. Luebker, and L. H. Wurm. Psychophysics of reading. XI. Comparing color contrast and luminance contrast. *Journal of the Optical Society of America A*, 7:2002–2010, Oct 1990.

[26] A. Capra, A. Castrorina, S. Corchs, F. Gasparini, and R. Schettini. Dynamic range optimization by local contrast correction and histogram image analysis. In *International Conference on Consumer Electronics (ICCE '06)*, pages 309 – 310, Las Vegas, Jan 2006.

Author Biography

__Marius Pedersen__ received his BsC in Computer Engineering in 2006, and MiT in Media Technology in 2007, both from Gjøvik University College, Norway. He is pursuing a PhD in Color Imaging, under the supervision of Pr. Hardeberg and Pr. Albregtsen, sponsored by Océ. He is also a member of the Norwegian Color Research Laboratory at Gjøvik University College. His work is centered on image quality metrics for color prints.