Attributes of image quality for color prints

Marius Pedersen
Gjøvik University College
Computer Science and Media Technology
P.O. Box 191
2802 Gjøvik, Norway
E-mail: marius.pedersen@hig.no

Nicolas Bonnier
Océ Print Logic Technologies
1 rue Jean Lemoine
94015 Creteil, France

Jon Yngve Hardeberg
Gjøvik University College
Computer Science and Media Technology
P.O. Box 191
2802 Gjøvik, Norway

Fritz Albregtsen
University of Oslo
P.O. Box 1080 Blindern
0316 Oslo, Norway

Abstract. *The evaluation of perceived image quality in color prints is a complex task due to its subjectivity and dimensionality. The perceived quality of an image is influenced by a number of different quality attributes. It is difficult and complicated to evaluate the influence of all attributes on overall image quality, and their influence on other attributes. Because of this difficulty, the most important attributes of a color image should be identified to achieve a more efficient and manageable evaluation of the image's quality. Based on a survey of the existing literature and a psychophysical experiment, we identify and categorize existing image quality attributes to propose a refined selection of meaningful ones for the evaluation of color prints.* © 2010 SPIE and IS&T. [DOI: 10.1117/1.3277145]

1 Introduction

Technology advancements are rapid in the printing industry. One goal of the printing industry, and a motivation for further advancements, is to produce high-quality prints both fast and economically. New and more refined ways to deal with the limitations of a printing system are proposed continuously to achieve high-quality prints. Quality assessment is needed to show whether technology advances improve the quality of a print.

There are basically two ways to judge image quality (IQ): subjectively or objectively. Subjective evaluation is carried out by observers and is therefore influenced by the human visual system (HVS). Objective evaluation of IQ can be carried out in many ways. One typical way is to use measurement devices to gather numerical values. Another way is to use algorithms, commonly known as IQ metrics, in an attempt to quantify image quality. IQ metrics are usually developed to take into account properties of the HVS, and thus have the goal of being well correlated with subjective evaluations.

Both objective and subjective evaluation of IQ are dependent on a number of quality attributes (QAs), which are terms of perception, such as sharpness, contrast, and saturation.[1] These QAs influence the overall IQ in different ways, and knowledge about their importance can be used to achieve an optimal reproduction of an image.[2] The importance of different QAs has been investigated and acknowledged by many researchers,[2–11] but, so far, there is no agreement on which QAs are the most important. In this paper we focus on QAs for the evaluation of color prints.

For many years, a goal in IQ evaluation has been to develop objective measures correlated with subjective quality. The advancements in this field have been driven by the desire to reduce the dependence on human observers and minimize both time and resources needed to quantify IQ. To achieve this, the QAs used for the subjective evaluation of quality need to be identified and their importance needs to be assessed. These objective measures can be used to

Paper 09066SSRR received May 1, 2009; revised manuscript received Nov. 3, 2009; accepted for publication Nov. 10, 2009; published online Jan. 13, 2010.

1017-9909/2010/19(1)/011016/13/$25.00 © 2010 SPIE and IS&T.

help observers detect quality issues, identify where loss of quality occurs in a printing workflow, or compare the quality of different printing systems.

IQ models have been created to establish a link between subjective and objective quality. These models are theories of perception that enable the prediction of IQ.[12] They are intended to describe the overall IQ of a system and to help researchers evaluate IQ. IQ models are composed of QAs, these models show how QAs relate to each other and their influence on the overall IQ. The goal for IQ models is to evaluate all QAs and their relationships—a very difficult and complex task. Most IQ models reduce the complexity by using a subset of the most important QAs. This subset can be defined based on technological issues, by asking observers, or by a combination of the two. By defining a subset of QAs, the strengths and weaknesses of a given system can be meaningfully represented with a relatively small number of QAs.[13] Several IQ models have been proposed,[3,7,14,15] but the search for better and improved IQ models continues.

This paper identifies and categorizes existing QAs to propose a refined selection of meaningful QAs for the evaluation of color prints. These QAs can be used to create a link between subjective and objective IQ in an IQ model. The identification and categorization of QAs can be used to assist in the evaluation of color prints and to improve or develop new evaluation methods, both objective and subjective.

This paper is organized as follows. Section 2 is a survey of QAs and IQ models. Section 3 discusses the selection of important QAs for the evaluation of color prints. Section 4 describes an experiment investigating a color workflow, where QAs are evaluated by observers. Section 5 concludes the discussion and suggests directions for further research in this field.

2 State of the Art

A brief survey of QAs and IQ models is given in this section.

2.1 *Image Quality Attributes*

Norberg *et al.* evaluated overall quality, as well as color rendition, sharpness, contrast, detail rendition in highlight and shadow areas, color shift, gloss, mottle, and print homogeneity in a comparison of digital and traditional print technologies.[5] In a study by Lindberg, 12 different QAs (overall quality, gamut, sharpness, contrast, tone quality, detail highlights, detail shadow, gloss level, gloss variation, color shift, patchiness, mottle, and ordered noise) were used to evaluate color prints.[4] Based on the evaluation performed by human observers, these 12 QAs were reduced to two orthogonal dimensions using factor analysis; one related to print mottle and one related to color gamut. These two dimensions accounted for almost all variation in the data set. Gast and Tse evaluated six different QAs, including blur, noise, banding, color rendition, tone reproduction, and printer type.[16] These QAs were evaluated in terms of preference. Additionally, several researchers investigated the importance of QAs such as sharpness,[9] contrast,[17] artifacts (for example, noise[8] and banding[10]), naturalness,[2] and color.[3,11,17–19]

Research on the combined influence of QAs has been carried out as well. In 1980, Sawyer investigated the influence of sharpness and graininess on perceived IQ as well as their combined influence.[6] Two years later, Bartleson investigated the combined influence of sharpness and graininess on color prints.[7] Both Sawyer and Bartleson showed results in which the worst QA tended to determine the quality, and a change in other QAs would not increase quality. In 1999 Natale-Hoffman *et al.* investigated the relationship between color rendition and micro-uniformity on preference.[13] This was considered by the authors as a step toward predicting preference without depending on human observers.

The identification of QAs has also been recognized as important for IQ metrics. Morovic and Sun based an IQ metric on perceptual QAs, where the QAs were determined based on answers from observers.[11] Lightness, hue, chromaticity, details, and contrast were found to be important. Only the first three, being the most important according to Morovic and Sun, were incorporated in the IQ metric (ΔI_{cm}). Later, Wang and Shang showed that defined QAs were beneficial for training IQ metrics.[20]

2.2 *Image Quality Models*

A framework for IQ models was proposed by Bartleson in 1982.[7] His approach was divided into three parts:

1. identification of important QAs
2. determination of relationships between scale values and objective measures
3. combination of QA scale values to predict overall IQ.

Bartleson used this framework to investigate the combined influence of sharpness and graininess on the quality of color prints. This framework has the advantage of representing strengths and weaknesses of a given system by a relatively small number of QAs. Because of this advantage and the framework's perceptual considerations, this framework was adopted by several researchers.[3,12,14] We also adopted this framework for this paper, where we mainly discuss the first part, identification of important QAs.

Dalal *et al.* followed Bartleson's framework to create the document appearance characterization system, which is a two-sided appearance-based system composed of QAs: one part for the printer and one for materials and stability.[3] For most QAs in the system, the evaluation is performed by experts. The basic IQ is given by a total of 10 QAs for both the printer and materials and stability. These describe different aspects of the system, including color rendition, uniformity, tone levels, and stability. The document appearance characterization system has several advantages. It uses high-level descriptors that cover a wide range of IQ issues, such as defects and sharpness. The printer is separated from materials and stability, allowing for separate analysis. The system also has a clear advantage by being technology independent. In addition, the QAs in the system are somewhat orthogonal (i.e., they do not influence each other).

The document appearance characterization system has some drawbacks as well. Since the evaluation is carried out mostly by experts, the results are influenced by the subjectivity of the expert. The system might be unsuitable for nonexperts due to its complexity, because the QAs are associated with known printing problems and technological issues. The approach of this system is different from the approach taken by Morovic and Sun,[11] where QAs were chosen based on answers from observers, resulting in more general QAs. The QAs in the model also are not adapted to IQ metrics, making it difficult to obtain a completely objective evaluation of IQ. However, the document appearance characterization system was not intended to use only IQ metrics, since it was made for subjective evaluation by experts. In addition, the system does not directly account for the contrast QA, which has been regarded as an important QA by other researchers.[4,5,11,17]

Keelan[14] also adopted the framework proposed by Bartleson.[7] He first identified important QAs, then found the relationship between a subjective scale (based on just noticeable differences) and an objective metric. In cases where multiple QAs influence the quality of an image, Keelan's approach found the influence of each QA to overall IQ. Keelan adopted multivariate formalism as a tool to combine the influence of each QA and obtain a value for overall IQ. QAs used in Keelan's model were assumed to be independent, which is different from the QAs used by others.[4,5,11] The advantage of Keelan's model is that QAs do not influence each other and can be easily combined to achieve an overall IQ, value which is not straightforward for dependent QAs. However, the disadvantage is that it might be very difficult to identify independent QAs. Keelan's model also assumes that objective metrics can be readily designed. Nonetheless, Keelan proposed a method to deal with this issue; rather than considering several QAs, the problem was approached by considering each nonindependent QA as a single QA with several facets.

Engeldrum focused on building an IQ model that partially adopted Bartleson's framework.[12] He proposed the IQ circle, which is based on four elements: customer quality preference, technology variables, physical image parameters, and customer perceptions. The last element, customer perceptions, contains the perceptual QAs (or "nesses"), which is the topic of this study. The IQ circle shows the relationship between objective and subjective quality, but it does not include which nesses are important nor how they should be quantified. Engeldrum also stated that observers most likely would not be able to perceive more than five QAs simultaneously. This statement is contradictory to the other IQ models that use QAs, such as the document appearance characterization system, in which a total of 20 QAs (10 for each side of the system) were evaluated. Norberg *et al.* evaluated overall IQ and nine QAs.[5] Lindberg evaluated overall IQ in addition to 12 QAs, but analysis showed that these QAs could be reduced to only two QAs.[4]

Many other IQ models have been proposed.[11,15,21] Some of these used IQ metrics, which calculate one or several values to represent IQ. IQ metrics can be full-reference, reduced-reference, or no-reference. The first type accepts an original and uses it to calculate IQ, the second type accepts parts of an original (for example, a subsignal) in the calculation, and the third type does not use information from an original. One of the full-reference metrics is S-CIELAB,[15] where spatial preprocessing of the image is carried out before the CIELAB color difference formula[22] is applied to calculate IQ. Metrics like S-CIELAB and others are most often constructed to quantify either overall quality or the quality of specific QAs. These metrics usually incorporate several stages of processing. Each stage is linked to a specific aspect of IQ, where characteristics of different QAs are taken into account. There are several different approaches to measure IQ. S-CIELAB[15] and ΔI_{cm}[11] are built on the idea that color differences are responsible for a large proportion of the differences between an original and its reproduction. Another IQ metric, structural similarity (SSIM), is based on the degradation of structural information.[27] Many of these metrics were proposed for different purposes, such as image difference and image fidelity. For a complete overview of full-reference IQ metrics, we refer the reader to Pedersen and Hardeberg.[23]

2.3 *Important Issues Regarding the Selection of QAs*

Our investigation of the existing literature on QAs and IQ models revealed several issues that must be dealt with in the selection of QAs and their use with IQ models. Usually tradeoffs exist among the different issues that require compromises.

The selection of QAs can be based on different aspects, such as technological issues or perception. Several aspects influence the selection of QAs, such as the basis on which to select the QAs and the intended use of the QAs. QAs based solely on technological issues might not be suitable to evaluate perceived IQ, and vice versa.

The basis upon which QAs have been selected also affects the evaluation of QAs, whether subjective or objective evaluation methods are used. For subjective evaluation, the complexity of the QAs determines the required expertise level of the observer. For objective evaluation, some QAs might be specially designed for measuring devices, while others are intended for IQ metrics.

As mentioned previously, IQ models usually work on a subset of QAs. The number of QAs, commonly referred to as the dimensionality, is important for the evaluation. According to Engeldrum, the use of more than five QAs is most likely unnecessary. A tradeoff exists between the precision of different QAs and the number of QAs used. Using a higher number of QAs will give a more precise view of the QAs and IQ; however, observers in a perceptual experiment might not consider all of the QAs, and QAs that are not considered will most likely not influence the perceived IQ. If all these QAs are considered, either by observers or objective methods they might provide inaccurate or in the worst case less precise results.

Another very important issue is independence. If the QAs are independent, the quality that results from them can easily be combined to obtain an overall IQ value. However, it is difficult to identify completely independent QAs, and special care must be taken when analyzing the results or combining the results from different QAs to obtain a value for overall IQ. Keelan proposed a method to test for interactions between QAs that used different sample series in perceptual experiment.[14]

The size (i.e., the number of sub-QAs or range of values) of the different QAs will also have impact, both on how they are used and how to analyze them. The skewness of QAs in an IQ model must be addressed, and it is not a straightforward process to combine QAs of different sizes.

Several key issues that must be dealt with when selecting QAs and building IQ models are summarized here:

- origin of QAs
- intended use
- dimensionality
- independence
- QA size.

These key issues will be discussed in the next section, where we investigate and select QAs for the evaluation of color prints.

3 Investigation and Selection of Important Quality Attributes

The first step in developing an IQ model is to identify the relevant and important QAs. We took the approach of doing a survey of the existing literature. Numerous QAs have been considered as important and evaluated by researchers to quantify IQ. To avoid excluding QAs in this part of the investigation, we included QAs based on both technology and perception, and QAs used with different intentions. These QAs include, lightness,[11,14] sharpness,[4,5,9,24,25] blur,[16] contrast,[4,5,11,17] banding,[10,16,29,30] contouring,[26] noise/graininess,[6–8,16,24,27,28] details,[5,11,17,18,24] naturalness,[2] color,[3,17,18] hue,[11,19] chroma,[11] saturation,[17] color rendition,[3,16] process color gamut,[3] artifacts,[17] mottle,[4,25] gloss,[4,5] color reproduction,[28] tone reproduction,[16,28] color shift,[5,25] ordered noise,[5] patchiness,[5] line quality,[3,31] text quality,[3] gamut size,[32] adjacency,[3] printer type,[16] effective resolution,[3] effective tone levels,[3] gloss uniformity,[3] skin color,[18] paper roughness,[25,33] paper flatness,[3] paper whiteness,[25,34] perceived gray value,[24] structure changes,[24] micro uniformity,[3] macro uniformity,[3] structure properties,[24] color gamut,[25] correctness of hue,[35] colorfulness proportional to the original,[35] correctness of lightness,[35] edge sharpness,[31] and edge raggedness.[31]

When reducing these QAs we literature, we considered several important issues, as mentioned previously, including the intended use of the QAs, and their origin. A long-term goal of this research is to create a link between subjective and objective IQ of color prints. With this intention, the QAs had to be based on perception and account for technological printing issues. The QAs had to be general enough to be evaluated by observers; and in order to not exclude novice observers, the QAs had to be somewhat straightforward to evaluate. In addition, the QAs had to be suitable for IQ metrics to address the intended objective method. The existing sets of QAs and models did not fulfill all of these requirements, and therefore a new set of QAs was needed.

Many of the QAs listed above are similar and have common denominators, which allows them to be grouped within more general QAs to reduce the dimensionality and create a more manageable evaluation of IQ. Usually a compromise is necessary between generality and accuracy when it comes to dimensionality. A small set of general QAs results in lower accuracy but also lower complexity, while a higher dimensionality offers accuracy but with greater complexity. We assigned most of the above QAs to six different dimensions considered important for the evaluation of IQ, which resulted in a reasonable compromise between accuracy and complexity. It is also a number close to that identified by Engeldrum,[12] who stated would

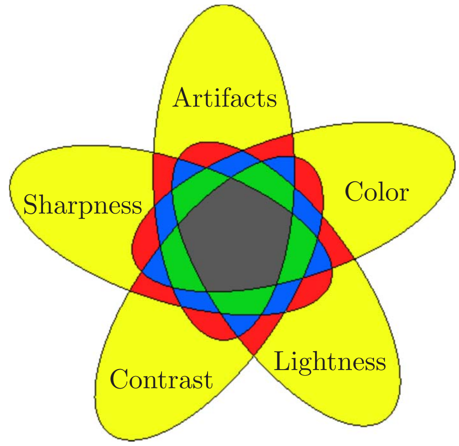

Fig. 1 Simple Venn ellipse diagram with five folds used for an abstract illustration of five different different QAs and the interactions between then. Overall IQ can be influenced by one (yellow), two (red), three (blue), four (green), or five (gray) of the QAs. (Color online only.)

that observers not perceive more than five QAs simultaneously. We reduced the QAs found in the literature to the following six:

1. The **color** QA contains aspects related to color such as hue, saturation, and color rendition, except lightness.
2. The **lightness** QA is considered so perceptually important that it is beneficial to separate it from the color QA.[14] Lightness ranges from “light” to “dark.”[1]
3. The **contrast** QA can be described as the perceived magnitude of visually meaningful differences, global and local, in lightness and chromaticity within the image.
4. The **sharpness** QA is related to the clarity of details[9] and definition of edges.[36,37]
5. The **artifacts** QA includes noise, contouring, and banding. In color printing, some artifacts can be perceived in the resulting image. These artifacts can degrade the quality of an image if they are detectable.[38,39]
6. The **physical** QA contains all physical parameters that affect quality, such as paper properties and gloss.

These six QAs are concise, yet comprehensive, high-level descriptors, being either artifactural (those which degrade the quality if detectable[38,39]) or preferential (those which are always visible in an image and have preferred positions[39]).

We used Venn diagrams to create simple and intuitive illustrations of the QAs and their influence on overall IQ. Venn diagrams may be used to show possible logical relations between a set of attributes. However, it is not possible to create a simple Venn diagram with a six-fold symmetry.[40] Therefore we illustrated the QAs using only five folds, leaving the physical QA out. This does not mean that the physical QA is less important than the other QAs.

The Venn diagram in Fig. 1 illustrates how the overall IQ is influenced by one, two, three, four, or five of the QAs. Many of the QAs are interdependent,[41] making quality a multidimensional issue,[42] in this case five dimensions. These QAs can influence the overall quality in different ways, and therefore the ellipses may not have equal sizes or the same positions in all situations. It is difficult to reduce all of the QAs used in the literature to six dimensions while preserving independence and including perceptual QAs. It is very difficult, if not impossible, to do this while accounting for most of the aspects of IQ for color prints. These difficulties present a disadvantage, since they lead to a more complex analysis of the results. However, methods have been proposed to deal with this problem.[14]

The six QAs identified above are a good starting point for the quality evaluation of color prints, and they can be adapted to different situations. Each of these QAs can be divided into sub-QAs for adaptation to specific issues, and thereby increase the accuracy of the QAs. For example, the artifact QA can be divided into three sub-QAs: noise, contouring, and banding. A separate analysis of these sub-QAs can be advantageous since it allows for specific analysis either by experts or IQ metrics. When there is a skewness in the distribution of QAs, separating the QAs into sub-QAs can also prove useful for improving the balance among the QAs. Furthermore, some sub-QAs, such as uniformity, can apply to several main QAs. This sub-QA can be placed under the color QA, but also under the artifacts QA, since a lack of uniformity can be thought of as an artifact. The placement of these sub-QAs must be done where it is most appropriate. Additionally, not all QAs might be used for a given evaluation of IQ. By excluding some QAs and dividing QAs into sub-QAs, we can consider the QAs used by other researchers as special cases of our proposed QAs.

Next we will take a closer look at the six different QAs and identify links to QAs from the literature.

3.1 *Color*

Color is a sensation. It is the result of the perception of light by the HVS.[43] The color QA includes color-related issues like hue, saturation, and color rendition. Lightness is not a part of our color QA. Since our HVS processes lightness and chromaticity information differently, it is convenient to treat these as separate QAs.[14]

Many of the QAs used in the literature can be linked to one of the six proposed QAs. Within our color QA, we can find several QAs used by other researchers, such as color,[3,17,18] hue,[11,19] chroma,[11] saturation,[17] color rendition,[3,16] process color gamut,[3] color reproduction,[28] color shift,[5,25] gamut size,[32] skin color,[18] color gamut,[25] correctness of hue,[35] and colorfulness proportional to the original.[35] Many of these can also be connected to the other QAs discussed below.

3.2 *Lightness*

A common definition of lightness is the visual sensation by which the area where the visual stimulus is presented appears to emit more or less light in proportion to that emitted by similarly illuminated areas perceived as a "white" stimulus.[1] Variations in lightness range from "light" to "dark."[1]

Many QAs used by other researchers can be included within our lightness QA, like tone reproduction,[16,28] perceived gray value,[24] correctness of lightness,[35] and lightness.[11,14] Other QAs are more difficult to link with just one QA, such as color shift,[5,25] gamut size,[32] color gamut,[25] process color gamut,[3] and color rendition.[3,16] All of these QAs have ties to the lightness QA, but also to the color QA. Other QAs will influence lightness but cannot be accounted for within the lightness QA, such as paper flatness and gloss level.

3.3 *Contrast*

Contrast is a difficult QA since there are many different definitions of contrast.[44–46] Michelson[47] defined contrast as

$$\frac{I_{\max}-I_{\min}}{I_{\max}+I_{\min}},$$

where $I_{\max}$ and $I_{\min}$ represent the highest and lowest luminance. In Weber,[48] contrast is defined as

$$\frac{I-I_b}{I_b},$$

where I represents the luminance of features, and I_b is the background luminance. Root-mean-square (RMS) contrast is another common way to define contrast:

$$RMS=\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^{1/2},$$

where x_i is a normalized gray level value, and $\bar{x}$ is the mean normalized gray level. Contrast can also be defined as the visual property that makes an object distinguishable. This definition is useful to express the readability of prints. Another definition of contrast is the lightness ratio between two areas in an image. Fedorovskaya defined contrast as an integrated impression of differences in lightness, or lightness variation observed within the whole picture.[37] Keelan defined contrast, in the context of color and tone reproduction, as the relationship between the original scene lightness perceived by the photographer and the final image (reproduced) lightness perceived by the observer.[14]

Contrast is clearly difficult to define, and its definition changes according to the application. Even so, the literature distinctly presents some common characteristics of contrast. This commonality is related to chromaticity,[49,50] and because of this, a definition of contrast based solely on lightness cannot describe perceived contrast in color prints. As well as being correlated with color, contrast is also related to local and global impressions.[49,50] Therefore, definitions that attempt to work on a global aspect are unsuitable for defining contrast in color prints.

The proposed QA is a perceptual QA, so a definition should relate to the HVS. To account for all the characteristics of contrast, we define contrast for color prints as the perceived magnitude of visually meaningful differences, both global and local, in the lightness and chromaticity within the image. This definition takes into account the characteristics of contrast and is applicable to color prints.

Our contrast QA can be linked with the use of contrast by several researchers.[4,5,11,17] Due to our definition of contrast, it will have ties with many different other QAs, such as chroma, saturation, and lightness.

3.4 *Sharpness*

Caviedes and Oberti related the perception of sharpness to the clarity of detail and edge definition of an image.[36] Bouzit and MacDonald followed a similar thinking and related sharpness to details and edges.[9] Fedorovskaya defined overall sharpness as the overall impression of clarity of edges observed within the whole picture.[37] Sharpness is related to both details and edges, and because of this our sharpness QA is defined as the clarity of details and edge definition of an image.

QAs that are suitable to group within the sharpness QA are diverse and many, including sharpness,[4,5,9,24,25] details,[5,11,17,18,24] line quality,[3,31] adjacency,[3] blur,[16] effective resolution,[3] edge sharpness,[31] and edge raggedness.[31]

3.5 *Artifacts*

Different printing artifacts are found in our artifacts QA, such as noise, contouring, and banding. It is common to all of these that if they are detectable, they contribute to degradation of the quality of an image.[39] In this QA, definitions of sub-QAs can be useful. For printing, noise, contouring, and banding are often considered to be important artifacts. Image noise can be defined as random variations in brightness or color in an image. Contouring can be characterized as the perceived color change within a small region exceeding a threshold,[51] which results in perceived discontinuity. Banding is the presence of extraneous lines or bands in a printed page,[10,29] that appear as nonuniform light or dark lines across the page.

We can link the proposed artifacts QA with several of the QAs used in the literature. For example, contouring,[26] noise/graininess,[6–8,16,24,27,28] banding,[10,16,29,30] artifacts,[17] effective tone levels,[3] mottle,[4,25] ordered noise,[5] patchiness,[5] structure changes,[24] and structure properties.[24] All of these will degrade quality if detectable.

3.6 *Physical*

Our proposed physical QA is important because the other five QAs cannot account for physical QAs such as paper roughness and gloss level. Our investigation of the literature revealed these physical QAs to be very important for the overall IQ, and therefore they should be accounted for in the evaluation of IQ. Several QAs used by researchers fit within this attribute, like paper roughness,[25,33] paper flatness,[3] gloss,[4,5] and printer type.[16]

3.7 *Relations Among Quality Attributes*

The six proposed QAs are not necessarily independent, and in order to calculate overall IQ, it is important to know which QAs influence other QAs and the magnitude of their influence. Our investigation of the literature revealed many of the relationships among QAs. Color can be linked to a number of other attributes, including contrast.[39,44] Color differences can also be linked to different artifacts, such as contouring[51] and banding.[52] One way to preserve details in gamut mapping is to introduce a slight color difference,[17] which relates color to the sharpness QA where details are included. Other relations to sharpness can be found as well; in the case of sharpening, a color difference can be introduced.[53,54] In extreme cases, sharpening can result in halo artifacts caused by color differences.[14]

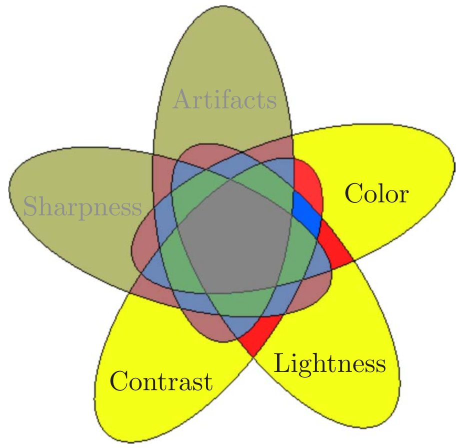

Fig. 2 Venn diagram showing a psychophysical experiment carried out to investigate QAs in a color workflow. In this experiment, a subset of the main QAs is considered to affect overall IQs color, lightness, and contrast. (Color online only.)

The last two links to the QA color can also occur due to changes in lightness, creating relations to sharpness and artifacts. Lightness can be linked to contrast,[37,44,55] but also to artifacts, such as banding[52] and contouring.[51]

Contrast has already been linked to color[39,44] and lightness.[37,44,55] It can be linked to sharpness as well,[55–61] since an increase in contrast generally increases sharpness.[58] In the literature we also find relations to contouring[62] and banding[63] artifacts.

Sharpness has also been linked with color,[17,53,54] artifacts,[14] and contrast.[55–61] Another artifact that is commonly mentioned regarding sharpness is noise, and the relationship between these QAs has been extensively examined in the literature.[38,57,58,64,65]

Artifacts can be related to a number of QAs. It has already been mentioned that noise relates to sharpness[38,57,58,64,65] and halos to lightness and color.[14] While contouring is linked to contrast,[62] banding can be related to both color and lightness, since the bands can be caused by lightness and or color variations,[52] but also to contrast.[63] The relations for artifacts will change according to the different artifacts evaluated.

Among the physical QAs, many relations can be found. For example, paper characteristics can influence color[34] and artifacts (as lack of smoothness[25]), while paper coating can affect artifacts (for example, lack of uniformity[5]).

In some situations, an increase in the quality of one QA might reduce the quality of another QA. One example can be the tradeoff between noise and sharpness, which has been investigated in the literature.[7]

The relations mentioned here do not address the magnitude of influence between QAs, which will be dealt with in future work.

4 Investigation of Quality Attributes in a Color Workflow

In this section we will take a closer look at the first three QAs: color, lightness, and contrast (see Fig. 2). We conducted on experiment to investigate quality issues in a color workflow and to confirm the QAs proposed in the previous section. A set of images was investigated by 15 observers to determine the most important QAs and which QAs the observers used in the quality evaluation of a color workflow. The observers were both male and female and ranged from experts to nonexperts.

The images were reproduced using the International Color Consortium (ICC) perceptual rendering intent, which adjusts color appearance to achieve the most attractive result on a medium different from the original.[66] In the evaluation of this color workflow, the observers evaluated different QAs, the influence that these QAs had on overall IQ, and how they affect the observers' judgment of quality. For some QAs, the quality decreased, for other QAs, the quality increased; and some QAs neither increased nor decreased quality. Since investigating only the QAs that influence quality might no give an incorrect representation of which QAs were important, we required the observers to investigate all the QAs. Thus, the instructions we gave to the observers were crucial for obtaining correct results.

4.1 *Experimental Setup*

4.1.1 *Test images*

To ensure the observers use a sufficiently large set of QAs, they were given a broad range of images, natural as well as test charts, so the experiment would include different quality issues.[67] To achieve this, we followed the recommendations of Field[68] and CIE,[69] who chose test images based on the following criteria:

- 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 as skin tones, grass, and sky blue
- color transitions
- neutral gray.

Most of the images were pictorial with a wide range of scenes such as landscapes, portraits, and personal items (jewelry, books, and clothes). This variety helped to characterize the impacts for QAs[14] and ensured that the observers examined a wide variety of QAs. In addition to the pictorial images, test charts were included that were content-free and included a selection of "interest area" colors suitable for evaluation of different aspects of IQ.[68]

The experiment included a total of 56 images, as shown in Fig. 4(b): seven images from ISO,[70] two images from CIE,[69] three test charts, and 44 other images captured by the authors. The 44 images captured by the authors were in RAW format and converted to RGB using Camera Raw 5.0 in Adobe PhotoShop CS4 with a resolution of 150 dpi and a 16-bit encoding. The RAW images were manipulated by a professional to look optimal, then verified by two other professionals.

The images were printed at a resolution of 150 pixels per inch and at a size resulting in a printed reproduction at approximately 8 by 10 inches.

4.1.2 *Color workflow*

The first step in the color workflow was to re-render the set of images from sRGB to the perceptual reference medium gamut[71] using the sRGB v4 perceptual transform. The output profile (Fig. 3) was generated using the TC3.5 CMYK test target, measured using a GretagMacbeth Eye-One Pro spectrophotometer and ProfileMaker Pro 5.0.8. As a second step, we applied linear lightness scaling compensation in the CIE *XYZ* color space plus the hue-preserving minimum ΔE clipping gamut mapping algorithm[69] to the images to re-render them from the perceptual reference medium gamut to the gamut of the printing system. The linear lightness scaling was made between the black point CIELAB coordinates of each image to the black point CIELAB coordinates of the printing system contained in the output profile. The last step was to convert the color data from the profile connection space values to the CMYK values of the printing system by a relative colorimetric transform. The images were then printed with the Océ Color-Wave 600 wide format CMYK printer on Océ Red Label paper.[*]

4.1.3 *Viewing conditions*

Some observers were presented with a reference image on an EIZO ColorEdge CG224 and some on an EIZO ColorEdge CG221 since the experiment was carried out in two locations. The reference image was displayed at a color temperature of 6500 K and a luminance level of 80 cd/m^2 in accordance with sRGB specifications. This set was rendered for sRGB display, so a monitor capable of displaying the sRGB gamut was the most adapted reproduction device for this set of images. In addition, the display was fitted with a monitor hood to prevent glare. The printed images were presented randomly in a controlled viewing room at a color temperature of 5200 K, 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 was set up to follow the CIE guidelines[69] as closely as possible.

4.1.4 *Instructions given to the observers*

The instructions given to the observers focused on the overall quality rating of the reproduction and an which QAs the observer used in their evaluation. The instructions specified that all QAs used in the evaluation should be stated, even if they did not influence IQ. The following two questions were given to the observers:

1. Is the printed image a pleasing reproduction?
2. According to you, which quality attributes influence the quality of the reproduction?

For the first question a scale from 1 to 7 was given to the observers where 1 represented the most pleasing reproduction. The following description for each level was provided to help the observers in their evaluation:

1. most pleasing you can imagine
2. highly pleasing

[*]Océ LFM 050 Red Label specifications: 75 g/m^2, whiteness CIE 159, thickness: 99 μm, acidity 7.5 pH, ISO brightness R457+UV 108%, ISO brightness R457—UV 88%, opacity 92%, not coated.

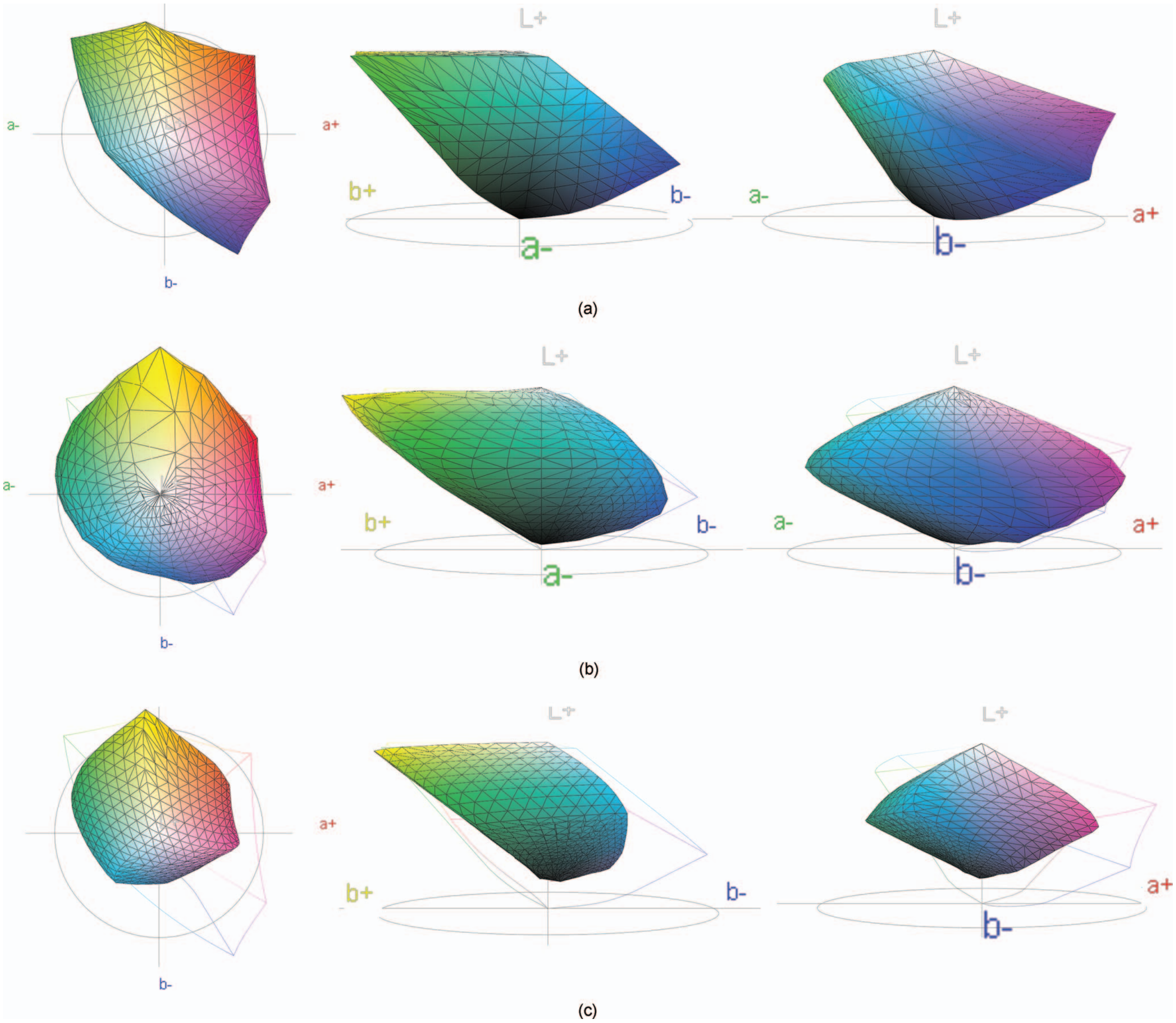

Fig. 3 Gamuts for the experiment shown with three different projections in the CIELAB space. The sRGB gamut is on the top, the perceptual reference medium gamut is in the middle, and the gamut of the printer is on the bottom. The small outer wireframe indicates the sRGB gamut boundary, and the circle indicates 100 on the *a* and *b* axis.

3. very pleasing
4. fairly pleasing
5. moderately pleasing
6. poorly pleasing
7. least pleasing reproduction possible.

The QAs used during the experiment were noted on a form, where QAs that decreased, increased, or did not influence quality were denoted with different symbols. Since we wanted to see how the QAs used by the observers fit within our proposed QAs, no descriptions or proposals for QAs were given to the observers to prevent any influence on the QAs and vocabulary used by the observers.

4.2 *Experimental Results*

Fifteen observers participated in the experiment. Five observers rated the whole data set, and 10 observers rated parts of the data set. A total of 452 evaluations were carried out by the observers, where a scale value was given to each image and the observers described the QAs they used in their evaluation. The evaluation was carried out in several sessions to prevent observer fatigue.

The overall average pleasantness of all images, based on the seven-level scale, was found to be between fairly and very pleasing [Fig. 4(a)]. The whole scale was used in the experiment. Ten observers used at least one of the extremes on the seven-level scale. An analysis of the observers' ratings indicates that images where a majority of the large areas had the same color (especially shadow areas) and images with color transitions (both test charts and natural images) were rated as least pleasing. An analysis of the observers' answers indicates that contouring was found in the images with color transitions, and since the contouring was highly perceivable in these images, the pleasantness

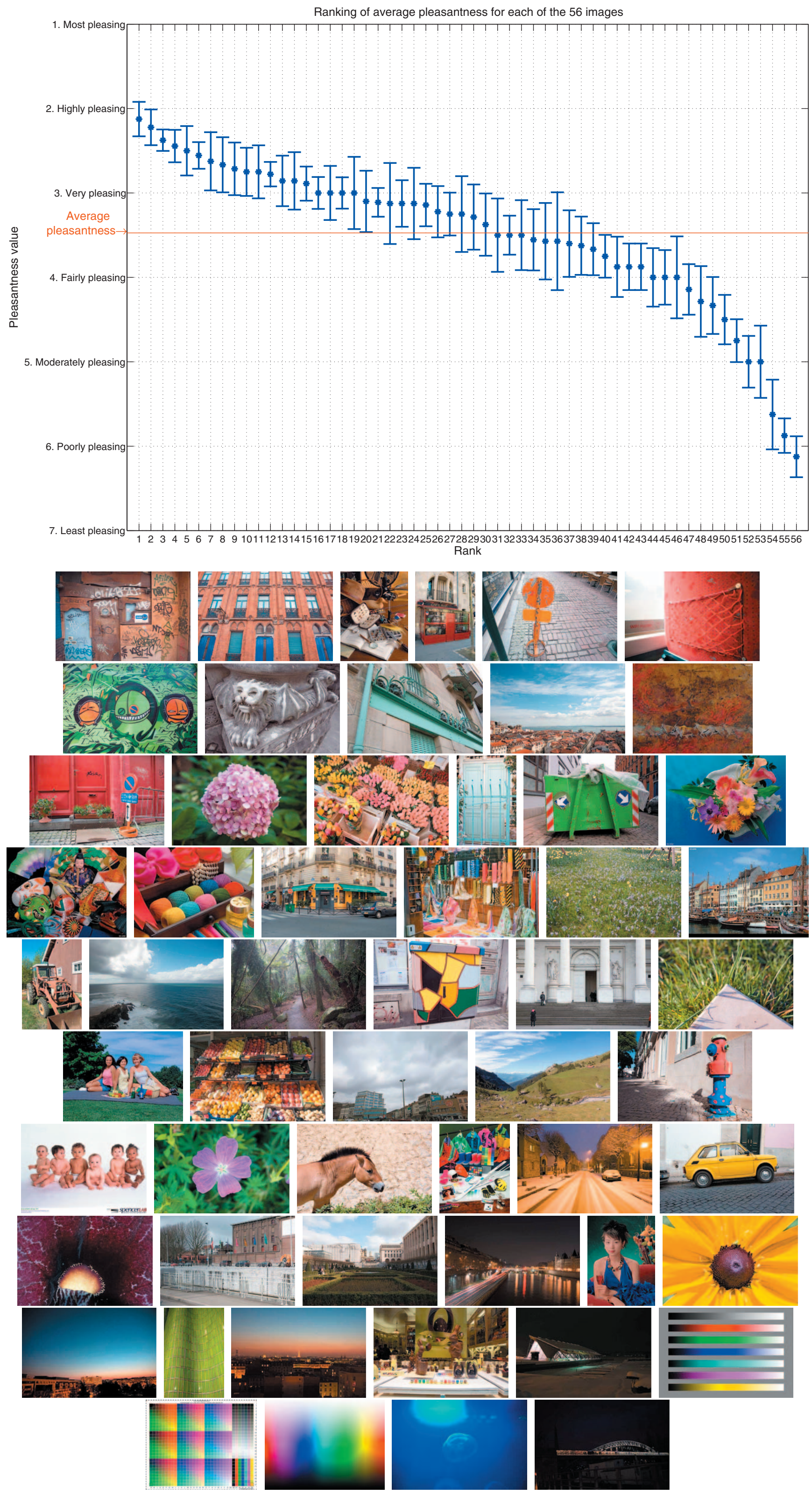

Fig. 4 (a) Average pleasantness rating for the 56 images in the experiment sorted from the most pleasing image to the least pleasing image. Each rating is plotted with a 95% confidence interval. (b) Thumbnails of the images below the graph are sorted in the same order, from left to right and top to bottom with the most pleasing image in the top left corner and the least pleasing image in the bottom right corner.

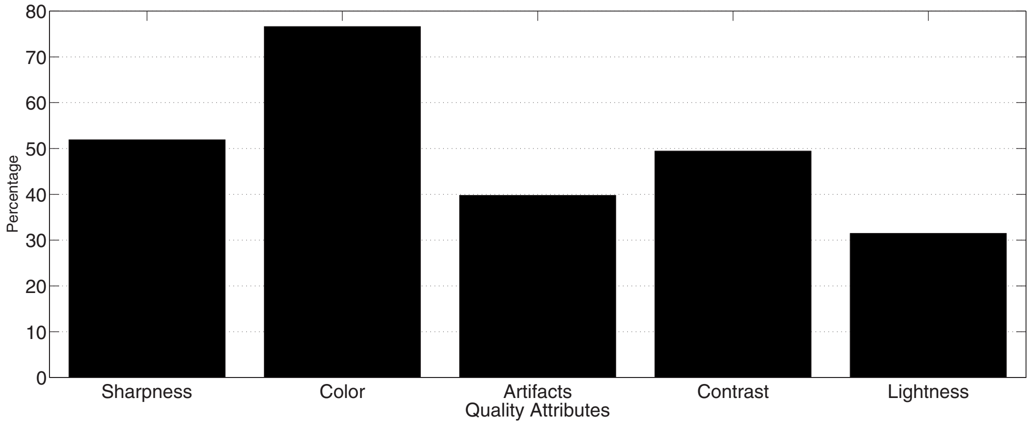

Fig. 5 Frequency of QAs used by the observers in the experiment. All QAs used by the observers were fitted to one of the important QAs proposed in the previous section.

rating was low. In some images with shadow areas, details were lost or were less visible, mainly due to the difference between the input and output gamut shown in Fig. 3. In images with color transitions, color breaks and loss of smoothness reduced the pleasantness of the image, which is a result of the gamut clipping algorithm. The images rated to be most pleasant had average saturation, lightness, and contrast. An analysis of the observers' results also reveals that the most pleasant images were equal or more colorful than the original and had equal or better contrast than the original. Here we will focus on the QAs used by the observers; however, an evaluation and comparison of the quality between the sRGB v4 perceptual transform and sRGB v2 transform was carried out elsewhere.[72]

For all observers, a total of more than 50 QAs were used in the evaluation, with an average of 10 different QAs for each observer. For each image, an average of 2.95 QAs were used, with a minimum of 1 and a maximum of 8. This indicates that the observers did not consider a large number of QAs in their evaluation of IQ. Many of the QAs used in the experiment overlapped, such as lightness, brightness, luminance, and darkness. All of these were connected to the lightness of the image and were grouped within the lightness QA. A similar grouping was done for other QAs to fit within the proposed set of QAs (Fig. 1).

Color was the most frequently used QA by the observers. As shown in Fig. 5, color was used to describe the IQ of more than 70% of the images. This is not surprising since it was a color workflow that we were investigating. Also, the color QA is fairly wide, containing sub-QAs like hue, saturation, and colorfulness. These three sub-QAs were used often by the observers, and often used together. The second most used QA, sharpness, contains mainly two sub-QAs: edges and details. Details, both in the highlights and shadows, were used frequently by the observers. Some observers also commented that a loss of contrast led to a loss of perceived sharpness, since edges and details were less prominent. This phenomenon has also been acknowledged in the literature.[14,55,56] The third most frequently used QA, contrast, is a narrower QA than color and sharpness. Because of this, it is interesting to note the observers' frequent use of contrast in the evaluation of color prints. It also confirms the need for a contrast QA in the evaluation of IQ.

In the test images, sub-QAs such as noise, contouring (lack of smoothness), banding, and so on could be perceived. The term "artifacts" or its sub-QAs were used in approximately 40% of the images. Lightness was considered in more than 30% of the images. Even though it was the least frequently used QA, some observers used the more general term "color" rather than separating lightness and chromaticity.

An analysis of the relations among QAs was carried out using cross-tabulation and chi-square tests. The null hypothesis H_0 was that there was no relationship between two QAs. The alternative hypothesis H_1 was that there was a relationship between two QAs. The p-values from this analysis are shown in Table 1. For some combinations of two QAs given a 5% significance level, H_0 was rejected in favor of H_1. This result indicates a dependence between color and lightness, but also between lightness and sharpness, contrast and artifacts, and artifacts and lightness. These results show that the observers' use of these QAs occurred simultaneously, but they do not show how the QAs affected each other or the overall IQ.

In the experiment, the observers distinguished between QAs that decreased IQ, did not influence IQ, or increased IQ. The observers marked more QAs as decreasing quality than the two other groups and more QAs increasing quality than QAs that did not influence IQ. Figure 6 shows the distribution of the three groups for each of the five QAs. For the artifacts attribute, some observers stated that the lack of artifacts increased quality. The observers did not consider artifacts where it did not influence IQ, indicating that artifacts were only considered when they were perceivable or not present in areas where observers expected to find artifacts. In the sharpness QA a lack of details in several of the reproductions contributed to decreased IQ. Therefore, we did not find an equal distribution between increasing and decreasing IQ as was observed for color, contrast, and lightness.

We investigated the dependence between image charac-

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

	Sharpness	Color	Artifacts	Contrast	Lightness
Sharpness	0	0.9976	0.1676	0.2392	0.0052
Color	0.9976	0	0.2194	0.7226	<0.0001
Artifacts	0.1676	0.2194	0	0.0089	0.0081
Contrast	0.2392	0.7226	0.0089	0	0.1408
Lightness	0.0052	<0.0001	0.0081	0.1408	0

teristics and the usage of the three quality levels (decreasing, not influencing, and increasing) for each QA. Cross-tabulation showed a dependence between the use of artifacts that decreased quality and images classified as test charts. This is not surprising since contouring and a lack of smoothness were most perceptible in these images. A dependence between fine details and increased quality due to contrast was found as well; this indicates that contrast is important for the perception of details. Low lightness and an increased IQ due to lightness also showed mutual dependence.

In this experiment, the physical QA was not considered and was therefore not a part of the analysis. Some QAs used by the observers were difficult to link with one of the five important QAs, such as naturalness and warmness. These QAs can be linked with changes in other QAs, such as color and lightness.

5 Conclusion and Future Directions

In this paper we identified and categorized existing QAs, and proposed a refined set of selection criteria of the most meaningful QAs for the evaluation of color prints. The number of QAs considered to be important in IQ evaluation was reduced to a set of six QAs: color, lightness, sharpness, contrast, physical, and artifacts. These QAs present a good starting point to describe overall IQ, and they can be considered as a step toward achieving a link between objective and subjective IQ.

A psychophysical experiment was carried out to evaluate a color workflow where QAs used by observers were recorded and analyzed. Results obtained from this experiment support the proposed set of QAs.

Future work includes the investigation of interactions between different QAs, locally and globally, and their influence on overall IQ. The incorporation of IQ metrics in a framework should be considered to achieve an objective evaluation of IQ. Comparing different output devices and using office documents as input are both considered to be relevant further steps in the investigation of QAs.

Acknowledgments

The authors would like to thank Nicolas Cardin for his assistance with the experiment and all the observers that participated in the experiment, and Alessandro Rizzi and Gabriele Simone for their enlightening discussions regarding contrast. M.P. has been enabled by Océ-Technologies B.V. to perform the research activities that underlie this document. This document has been written in a personal

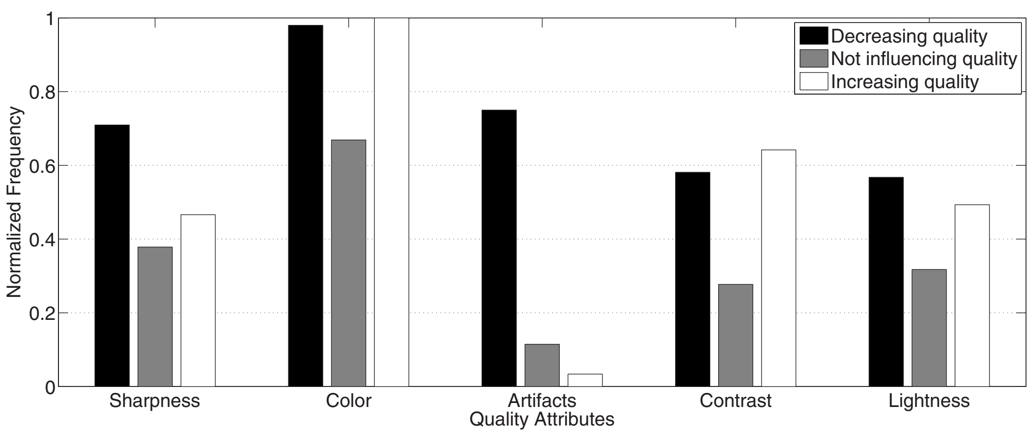

Fig. 6 Normalized frequency distribution of QAs that increase, decrease, and do not influence IQ.

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*, 2nd edition, Wiley Interscience, Derby, UK (2000).
2. E. A. Fedorovskaya, F. J. J. Blommaert, and H. de Ridder, "Perceptual quality of color images of natural scenes transformed in CIELUV color space," in *Color Imaging Conf.*, IS&T/SID, pp. 37–40 (1993).
3. E. N. Dalal, D. R. Rasmussen, F. Nakaya, P. A. Crean, and M. Sato, "Evaluating the overall image quality of hardcopy output," in *Image Processing, Image Quality, Image Capture, Systems Conf.*, IS&T, pp. 169–173 (1998).
4. S. Lindberg, "Perceptual determinants of print quality," PhD thesis, Stockholm University (2004).
5. O. Norberg, P. Westin, S. Lindberg, M. Klaman, and L. Eidenvall, "A comparison of print quality between digital, offset and flexographic printing presses performed on different paper qualities," in *International Conf. on Digital Production Printing and Industrial Applications*, IS&T, pp. 380–385 (2001).
6. J. F. Sawyer, "Effect of graininess and sharpness on perceived print quality," in *Photographic Image Quality Symposium*, Royal Photographic Society, pp. 222–231 (1980).
7. C. J. Bartleson, "The combined influence of sharpness and graininess on the quality of colour prints," *J. Photogr. Sci.* **30**, 33–38 (1982).
8. T. Sato, N. Aoki, and H. Kobayashi, "Improvement of digital print in preference by adding noise," in *NIP21: International Conf. on Digital Printing Technologies*, pp. 122–123, IS&T (2005).
9. S. Bouzit and L. MacDonald, "Colour difference metrics and image sharpness," in *8th Color Imaging Conference*, pp. 262–267, IS&T/SID (2000).
10. C. Cui, D. Cao, and S. Love, "Measuring visual threshold of inkjet banding," in *Image Processing, Image Quality, Image Capture, Systems Conf.*, pp. 84–89, IS&T (2001).
11. J. Morovic and P. Sun, "Visual differences in colour reproduction and their colorimetric correlates," in *10th Color Imaging Conf.*, pp. 292–297, IS&T/SID (2002).
12. P. G. Engeldrum, "Image quality modeling: Where are we?" in *Image Processing, Image Quality, Image Capture, Systems Conf.*, pp. 251–255, IS&T (1999).
13. K. Natale-Hoffman, E. Dalal, R. Rasmussen, and M. Sato, "Effect of selected image quality on overall preference," in *Image Processing, Image Quality, Image Capture, Systems Conference*, pp. 266–269, IS&T (1999).
14. B. W. Keelan, *Handbook of Image Quality: Characterization and Prediction*, Marcel Dekker, New York (2002).
15. X. Zhang and B. A. Wandell, "A spatial extension of CIELAB for digital color image reproduction," in *Soc. Inform. Display 96 Digest*, San Diego, pp. 731–734, IS&T/SID (1996).
16. G. Gast and M.-K. Tse, "A report on a subjective print quality survey conducted at NIP16," in *NIP17: International Conf. on Digital Printing Technologies*, pp. 723–727, IS&T (2001).
17. N. Bonnier, F. Schmitt, H. Brettel, and S. Berche, "Evaluation of spatial gamut mapping algorithms," in *14th Color Imaging Conf.* Vol. 14, pp. 56–61, IS&T/SID (2006).
18. J. Y. Hardeberg, E. Bando, and M. Pedersen, "Evaluating colour image difference metrics for gamut-mapped images," *Coloration Technol.* **124**(4), 243–253 (2008).
19. G. Hong and M. R. Luo, "New algorithm for calculating perceived colour difference of images," *Imaging Sci. J.* **54**(2), 86–91 (2006).
20. Z. Wang and X. Shang, "Spatial pooling strategies for perceptial image quality assessment," in *IEEE International Conf. on Image Process.*, pp. 2945–2948 (2006).
21. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
22. Commission Internationale de l'E'clairage (CIE). *Colorimetry*, 2nd ed., Publication CIE 15.2, Bureau Central de la CIE (1986).
23. M. Pedersen and J. Y. Hardeberg, *Survey of Full-Reference Image Quality Metrics*, Høgskolen i Gjøviks rapportserie 5, The Norwegian Color Research Laboratory, Gjøvik University College, (2009). ISSN: 1890-520X.
24. F. Nilsson and B. Kruse, "Objective quality measures of halftones images," in *NIP 13: International Conf. on Digital Printing*, pp. 353–357, IS&T (1997).
25. M. Klaman, "Aspects on colour rendering, colour prediction and colour control in printed media," PhD thesis, Stockholm University (2002).
26. P. Green, "Gamut mapping and appearance models in graphic arts colour management," PhD thesis, University of Derby (2003).
27. P. Oittinen, I. Huttu, and A. Mensonen, "End user preferences for quality of on-demand printed products," in *International Conf. on Digital Printing Technologies*, IS&T, pp. 147–152 (1996).
28. K. Miyata, N. Tsumura, H. Haneishi, and Y. Miyake, "Subjective image quality for multi-level error diffusion and its objective evaluation method," *J. Imaging Sci. Technol.* **43**(2), 170–177 (1999).
29. Y. Bang, Z. Pizlo, and J. P. Allebach, "Perception based hardcopy banding metric," in *NIP21: International Conf. on Digital Printing Technologies*, pp. 78–83, IS&T (2005).
30. Y. Bang, Z. Pizlo, and J. P. Allebach, "Discrimination based banding assessment," in *NIP19: International Conf. on Digital Printing Technologies*, pp. 745–750, IS&T (2003).
31. M.-K. Tse and A. H. Klein, "Automated print quality analysis in inkjet printing: Case study using commercially available media," in *Recent Progress in Ink Jet Technologies II*, pp. 515–519, Society for Imaging Science, Springfield, VA (1999).
32. M. Andersson and O. Norberg, "Color gamut: Is size the only thing that matters?" in *2006 TAGA Conf.*, Vancouver, Canada, pp. 273–291, Technical Association of the Graphic Arts (2006).
33. R. Xu, P. D. Flemming, A. Pekarovicova, and V. Bliznyuk, "The effect of ink jet paper roughness on print gloss," *J. Imaging Sci. Technol.* **49**(2), 660–666 (2005).
34. M. Klaman, "The influence of paper whiteness and other parameters on the creating of ICC-profiles for digital colour printers and conventional offset presses," in *IARIGAI: Advances in Printing Science and Technology—Advances in Digital Printing*, Vol. 6, Munich, Germany (1999).
35. R. W. G. Hunt, *The Reproduction of Color in Photography, Printing and Television*, Fountain, Tolsworth, UK (1987).
36. J. Caviedes and F. Oberti, "A new sharpness metric based on local kurtosis, edge and energy information," *Signal Process. Image Commun.* **19**, 147–161 (2004).
37. E. A. Fedorovskaya, "Image quality as a problem of computational vision," in *Image Processing, Image Quality, Image Capture, Systems Conf.*, IS&T, pp. 22–28 (2003).
38. N. Burningham, Z. Pizlo, and J. P. Allebach, "Image quality metrics," *Encyclopedia of Imaging Science and Technology*, pp. 598–616, Wiley, New York (2002).
39. K. Topfer, B. Keelan, S. O'Dell, and R. Cookingham, "Preference in image quality modelling," in *Image Processing, Image Quality, Image Capture, Systems Conf.*, pp. 60–64, IS&T (2002).
40. B. Grünbaum, "The search for symmetric Venn diagrams," *Geombinatorics* **8**, 104–109 (1999).
41. T. N. Pappas and D. L. Neuhoff, "Least-squares model-based halftoning," *IEEE Trans. Image Process.* **8**(8), 1102–1116 (1999).
42. F. Nilsson, "Objective quality measures for halftoned images," *J. Opt. Soc. Am. A* **16**, 2151–2162 (1999).
43. J. S. Friedman, *History of Color Photography*, American Photographic Publishing, Boston (1947).
44. M. Pedersen, A. Rizzi, J. Y. Hardeberg, and G. Simone, "Evaluation of contrast measures in relation to observers perceived contrast," in *CGIV 2008—Fourth European Conf. on Color in Graphics, Imaging and Vision*, pp. 253–256, IS&T (2008).
45. A. Rizzi, G. Simone, and R. Cordone, "A modified algorithm for perceived contrast in digital images," in *CGIV 2008—Fourth European Conf. on Color in Graphics, Imaging and Vision*, pp. 249–252, IS&T (2008).
46. E. Peli, "Contrast in complex images," *J. Opt. Soc. Am. A* **7**, 2032–2040 (1990).
47. A. Michelson, *Studies in Optics*, University of Chicago Press (1927).
48. P. Whittle, "The psychophysics of contrast brightness," in *Lightness, Brightness, and Transparency*, pp. 35–110, Lawrence Erlbaum Associates, Hillsdale, NJ (1994).
49. G. Simone, M. Pedersen, J. Y. Hardeberg, and I. Farup, "On the use of gaze information and saliency maps for measuring perceptual contrast," in *16th Scandinavian Conf. on Image Analysis*, A.-B. Salberg, J. Y. Hardeberg, and R. Jenssen, Eds., *Lect. Notes Comput. Sci.* **5575**, 597–606 (2009).
50. A. Tremeau, R. Nicolas, and E. Dinet, "A measure of colour contrast correlated with human perception," in *CGIV 2008 Fourth European Conf. on Color in Graphics, Imaging and Vision*, pp. 120–125, IS&T (2008).
51. C. Kuo, Y. Ng, and D. A. Gusev, "Perceptual color contouring detection and quality evaluation using scanners," in *NIP19: 2003 International Conf. on Digital Printing Technologies*, pp. 581–585, IS&T (2003).
52. R. Chung and M. Rees, "A survey of digital offset print quality issues," technical report, Rochester Institute of Technology (2007).
53. S. Y. Choi, M. R. Luo, and M. R. Pointer, "Modeling image naturalness," in *CGIV 2008—Fourth European Conf. on Color in Graphics, Imaging and Vision*, pp. 126–131, IS&T (2008).
54. J. Radun, T. Leisti, J. Häkkinen, H. Ojanen, J.-L. Olives, T. Vuori, and G. Nyman, "Content and quality: Interpretation-based estimation of image quality," *ACM Trans. Appl. Percept* **4**(4), 1–15 (2008).
55. A. J. Calabria and M. D. Fairchild, "Perceived image contrast and observer preference: I. The effects of lightness, chroma, and sharpness manipulations on contrast perception," *J. Imaging Sci. Technol.* **47**, 479–493 (2003).

56. A. J. Calabria and M. D. Fairchild, "Perceived image contrast and observer preference: II. Empirical modeling of perceived image contrast and observer preference data," *J. Imaging Sci. Technol.* **47**, 494–508 (2003).
57. G. M. Johnson, "Measuring images: differences, quality and appearance," PhD thesis, Rochester Institute of Technology (2003).
58. G. M. Johnson and M. D. Fairchild, "Sharpness rules," in *8th Color Imaging Conf.*, pp. 24–30, IS&T/SID (2000).
59. T. J. W. M. Janssen, "Computational image quality," PhD thesis, Technische Universiteit Eindhoven (1999).
60. D. Shaked and I. Tastl, "Sharpness measure: Towards automatic image enhancement," in *IEEE International Conf. on Image Processing (ICIP)*, Vol. 1, pp. 937–940 (2005).
61. P. G. J. Barten, "Evaluation of subjective image quality with the square-root integral method," *J. Opt. Soc. Am. A* **7**, 2024–2031 (1990).
62. L. Velho and J. Gomes, "Color halftoning with stochastic dithering and adaptive clustering," in *First European Conf. on Color in Graphics, Image and Vision*, pp. 627–632, IS&T (2002).
63. J. C. Briggs, M. Murphy, and Y. Pan, "Banding characterization for inkjet printing," in *Image Processing, Image Quality, Image Capture, Systems Conf.*, pp. 84–88, IS&T (2000).
64. G. Sharma, *Digital Color Imaging Handbook*, CRC Press, Inc., Boca Raton, FL (2002).
65. J. Dijk, "In search of an objective measure for the perceptual quality of printed images," PhD thesis, Technische Universiteit Delft (2004).
66. ICC, "White paper number 2: Perceptual rendering intent use case issues," Jan. 2005, ⟨www.color.org⟩ (16 March 2009).
67. J. Holm, J. Tastl, and T. Johnson, "Definition & use of the ISO 12640-3 reference color gamut," in *14th Color Imaging Conf.*, pp. 62–68, IS&T/SID (2006).
68. G. G. Field, "Test image design guidelines for color quality evaluation," in *The Seventh Color Imaging Conf.: Color Science, Systems and Applications*, pp. 194–196, IS&T (1999).
69. CIE, "Guidelines for the evaluation of gamut mapping algorithms," technical report CIE TC8-03, ISBN: 3-901-906-26-6, p. 156 (2004).
70. ISO, "Graphic technology—prepress digital data exchange. Part 2: XYZ/sRGB encoded standard colour image data (XYZ/SCID)," ISO 12640-2 (2004).
71. ISO, "Graphic technology—prepress digital data exchange. Part 3: CIELAB standard colour image data (CIELAB/SCID)," ISO 12640-3 (2007).
72. N. Bonnier, N. Cardin, C. Leynadier, and A. Sarlat, "Evaluating the use of the perceptual reference medium gamut in ICC printing works," in *Seventeenth Color Imaging Conf.*, IS&T and SID (2009).

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 currently 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.

Nicolas Bonnier is a color scientist with Océ Print Logic Technologies, which he represents in the International Color Consortium. He is also a part-time lecturer at the ENS Louis Lumiére (Paris), from which he graduated in 2000 with a major in photography. He received his masters degree in electronic imaging from Université Pierre et Marie Curie (Paris) in 2001. He was a member of the Laboratory for Computational Vision with Pr. Simoncelli at New York University from 2002 to 2005. He completed a PhD program in 2008 under the direction of Pr. Schmitt at Télécom Paristech, sponsored by Océ.

Jon Yngve Hardeberg is a professor of color imaging at Gjøvik University College. He received his PhD from Ecole Nationale Suprieure des Télécommunications in Paris, France, in 1999, with a dissertation on colour image acquisition and reproduction using both colorimetric and multispectral approaches. He has more than 10 years' experience with industrial and academic color imaging research and development and has co-authored over 100 research papers within the field. His research interests include various topics of color imaging science and technology, such as device characterization, gamut visualization and mapping, image quality, and multispectral image acquisition and reproduction. He is a member of IS&T, SPIE, and the Norwegian representative to CIE Division 8. He has been with Gjøvik University College since 2001 and is currently head of the Norwegian Color Research Laboratory.

Fritz Albregtsen received a Candidatus Realium from the Institute of Theoretical Astrophysics, University of Oslo, in 1978. He has been with the Department of Informatics, University of Oslo, since 1983, and professor of digital image processing since 1997. His research interests include automatic segmentation and analysis of medical images with a focus on statistical texture analysis for cancer diagnostics and prognostics, generating low-dimensional statistical features from very high-dimensional descriptions of information. He has a keen interest in metrics for the evaluation of quality of each step from imaging to the results of image processing and image analysis. He has coauthored two textbooks and about 135 refereed book chapters, journal papers, and publications in refereed international conferences, spanning from theoretical astrophysics to medical image processing and analysis.