Framework for the Evaluation of Color Prints Using Image Quality Metrics

Marius Pedersen,** and Seyed Ali Amirshahi*.*
** Gjøvik University College (Norway). ** Océ Print Logic Technologies S.A. (France).*

Abstract

Measuring the perceived quality of printed images is important to assess the performance of printers and to evaluate technology advancements. Image quality metrics have been proposed to objectively assess the quality of images, and new metrics are proposed continuously. However, since these metrics require digital inputs, applying these metrics to printed images are not straightforward. In order to accomplish this, the printed reproduction needs to be transformed into a digital copy.

In this paper we propose a framework for applying image quality metrics to printed images, including the transformation to a digital format, image registration, and the application of image quality metrics. The proposed framework introduces less error and is significantly faster than another state of the art framework. Finally, the framework is used to evaluate a set of image quality metrics against subjective data.

Introduction

When we print a digital image we get a physical copy of it, this copy differs from the digital original. The reproduction can be subject to a loss of Image Quality (IQ) due to the limitations of the printing system. One way to assess loss of quality is by using subjective evaluation. However, subjective evaluation is often time-consuming, inconvenient, resource demanding, and even expensive. In addition, human observers are not objective, and their preference of IQ may change over time. Objective evaluation of IQ can be used to avoid subjectivity and decrease other drawbacks of subjective evaluation. Many methods for objective IQ evaluation have been proposed, one of these is commonly referred to as IQ metrics [1], and their goal is to automatically predict IQ.

Subjective assessment of print quality is rather straightforward, where a group of observers can be asked about the quality of the printed image. However, assessment of a printed image by IQ metrics is not straightforward. The original image is of a digital format and the printed image is of an analog format, because of this the printed image must be digitalized before IQ assessment with IQ metrics. In this paper we discuss the transformation from a physical reproduction to a digital reproduction with the goal of proposing a framework for using IQ metrics to evaluate the quality of color prints.

The paper is organized as follows: first state of the art, then a framework based on control points is introduced, followed by application of the framework with IQ metrics, at last we conclude.

State of the Art

A few frameworks have been proposed for using IQ metrics on printed images. All these frameworks follow the same procedure; as a first step the printed image is scanned, which can be followed by a descreening procedure to remove halftoning patterns. Then image registration is performed to match the scanned image with the original. Finally, an IQ metric can be applied.

The first framework was proposed by Zhang et al. [2] in 1997. First, the image is scanned, then three additional scans are performed, each with a different color filter. This results in enough information to transform the images correctly to CIEXYZ. The printed image was scanned with an Agfa Horizon, and the scanning resolution was set to 1200 dpi. In this article no information about the image registration was given, nor on the descreening procedure. The applied IQ metric was S-CIELAB [3], and it was applied to color patches.

Another framework was proposed by Yanfang et al. [4] in 2008. Two control points were applied to the image before printing, one point to the upper left corner and one to the upper center, to help in the registration. The images were scanned with a Kodak i700 at 300 dpi before registration, where the two control points were used for matching the printed image with the original (Figure 1). Descreening was performed by the scanner at 230 lpi. No information was given regarding the scaling of the image. The applied IQ metric was S-CIELAB [3].

Recently, Eerola et al. [5] proposed a new framework (Figure 2), which follows the same steps of the previous frameworks. The printed reproduction is scanned, then both the original and the reproduction go through a descreening procedure, which is performed using a Gaussian low-pass filter. Further, image registration is carried out, where local features are used with a Scale-Invariant Feature Transform (SIFT). A RAnd SAmple Consensus principle (RANSAC) was used to find the best homography. This is different from the previous framework, since it uses local features instead of control points. Scaling was performed using bicubic interpolation, and the scanner resolution was set to 1250 dpi. LABMSE was the applied IQ metric.

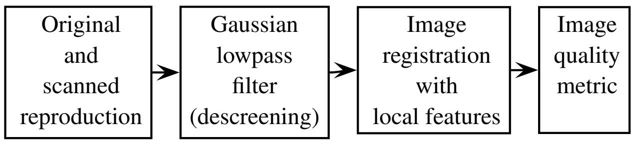

Figure 2. Framework from Eerola et al. [5], where local features are used rather than control points.

A Framework Based on Control Points

We modify and propose a framework similar to the framework by Yanfang et al. [4], which performs image registration based on control points. These control points act as the basis for the image registration, where the control points are used to perform different transformation procedures. As a first step in our

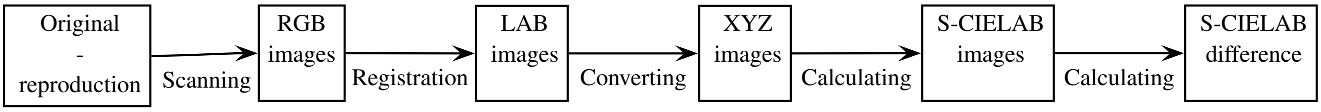

Figure 1. *Framework by Yanfang et al. [4]*

framework, the image is padded with a white border and equipped with four control points prior to printing. These points are black squares placed just outside the four corners of the image. Then the image is printed with a given device, such as an ink jet printer. Further, the image is scanned, and the profile of the scanner is assigned to the scanned image in order to achieve a correct description of the colors. The next step in the framework is to find the coordinates of the center of the control points, in both the original image and the scanned image. This is done by a simple automatic routine based on detection of squares. The scanned image can be affected by several geometric distortions, such as translation, scaling, and rotation. Because of this, image registration must be carried out. The coordinates for the control points, in both images, are used to create a transformation for the registration. There are several possible transformation types for doing the registration, which are discussed below. In addition, the interpolation method for scaling also has several possible methods, which are compared below. After the scanned image has been registered to match the original, a simple procedure is applied to remove the white padding and the control points. Finally, an IQ metric can be used to calculate the quality of the printed image. An overview of the framework is shown in Figure 3. This framework differs from the one from Eerola et al. [5], not only in the registration method, but also in the descreening. In our modified framework we do not perform any direct descreening, but we leave this to the IQ metrics in order to avoid a double filtering of the image.

Scanner

The first step for using the framework is to scan the printed image. There are two different scanner types available; Charge Coupled Device (CCD) and Contact Image Sensor (CIS). The disadvantage with an CIS scanner is the depth-of-view. However, this is not a problem with prints, since the print is usually laying flat on the scan surface. CIS scanner also has a significantly lower gamut and resolution than CCD scanner, but CIS has a stable illumination system. Due to the resolution and gamut issues we have used a CCD scanner. It is important that the scanner is able to cover the gamut of the printer.

Scanning Resolution

A central issue while scanning is to have a resolution high enough to capture the perceived details of the printed image. For the evaluation of the resolution a Microtek ScanMaker 9800XL scanner was used. The scanner was calibrated using a Kodak IT8.7 Target and the profile was created with ProfileMaker 5.0.8.

An image, containing both uniform areas and details, was prepared with control points just out side the corners. Then it was printed with an Océ ColorWave 600 wide format printer on Océ Red Label paper at a resolution of 150 pixels per inch. The printed image was then scanned with the following resolutions: 72, 100, 150, 300, and 600 dpi without any automatic corrections. The scanned image was used as input to the framework, where the errors were investigated using MSE and ΔE^*_{ab}. From the results in Table 1 we see that the values fluctuate, and does not give any indication of the resolution needed. Since the printed image will be different from the original, just finding the lowest objective value is not appropriate, because of this further investigation is needed.

Resolution (dpi)	MSE	ΔE^*_{ab}
72	7.5	14.5
100	6.5	12.5
150	5.7	12.1
300	5.9	12.9
600	6.1	14.6

Table 1. Scanning resolution test. MSE is in the 10^4.

The next step after the objective evaluation of the scanning resolution is subjective evaluation. Visual inspection of the images reveals a higher detail level in the 600 dpi scans, than at lower resolutions. However, at a normal viewing distance these details are not apparent in the printed image.

We have also downsampled the images scanned at different resolutions to match the original at 150 dpi, and visually compared the results over the different resolutions. We have compared 150 dpi and higher, because it does not make sense to look at lower resolutions since these must be upsampled. The visual inspection shows that 300 dpi and higher is smoother than 150 dpi, and that the halftoning pattern has been blurred out, due to the interpolation. This indicates that the downsampling procedure acts as a descreening. However, pixel-wise IQ metrics that does not take into account the Human Visual System (HVS) should be used with caution. In addition, it has been found that IQ metrics simulating the HVS are better than IQ metrics without simulation of the HVS [6].

The scanner resolution is also dependent on the visual angle the prints are evaluated and on the IQ metrics. In addition to our findings, Lim and Mani [7] state that a resolution of 600 dpi should be sufficient to capture the details of a printed image.

Image registration

When the image has been scanned it needs to be registered to match the original image, and there are several issues to consider in the registration process. First the transformation type, then the interpolation method.

Transformation Type

The transformation type to be used with the framework must use the same or fewer number of control points added to the image. In order to find the most appropriate transformation, we have tested five different transformation types:

- Nonreflective similarity uses two pair of control points, and is suitable when the image is distorted by some combination of translation, rotation, and scaling.

$$[\; u \quad v \;] = [x \quad y \quad 1]T, \tag{1}$$

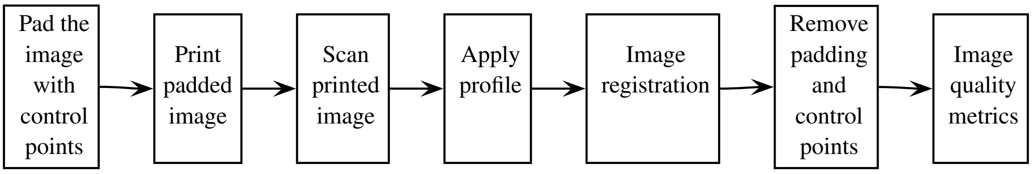

Figure 3. *Overview of the proposed framework for using IQ metrics with printed images.*

where x and y are the dimensions to be transformed, T is a 3-by-2 matrix defined as

$$\begin{bmatrix} sc & -ss \\ ss & sc \\ tx & ty \end{bmatrix},$$

where

$$sc = s * cos(theta)$$

and

$$ss = s * sin(theta).$$

s is the scale factor and $theta$ the rotation angle, tx and ty indicate the translation. u and v are the transformed dimensions.

- Similarity is resembling the previous, but uses reflection in addition. In this case T is defined as

$$\begin{bmatrix} sc & -a \times -ss \\ ss & a \times sc \\ tx & ty \end{bmatrix},$$

if $a = -1$ reflection is included.

- Affine uses three pairs of control points, and is useful when the image exhibit shearing.

$$[\ u \quad v\ ] = [x \quad y \quad 1]T, \tag{2}$$

where T is a 3-by-2 matrix where all six elements can be different, and at least three control-point pairs are needed to solve for the six unknown coefficients.

- Projective uses four pair of control points, and is commonly used when the images appears tilted.

$$[\ \frac{up}{wp} \quad \frac{vp}{wp} \quad w\ ] = [x \quad y \quad w]T, \tag{3}$$

where T is a 3-by-3, where all nine elements can be different, and t least four control-point pairs are needed to solve for the nine unknown coefficients.

- Piecewise linear uses four pair of control points, and is used when parts of the image appear distorted differently. This transformation method applies affine transformations separately to triangular regions of the image [8].

Theoretically correction of translation, rotation, and scaling are needed to register a scanned image. However, we wanted to test other transformation methods, such as to account for shearing and local distortions, which might occur in scanners as well. In order to find the most appropriate transformation type several tests have been performed, which are discussed below.

Interpolation Method

The interpolation method used for scaling the images are important, there are three main methods for doing this:

- bicubic interpolation,
- bilinear interpolation,
- nearest-neighbor interpolation.

The first method considers a 4×4 neighborhood, and accounts for the distance to the unknown pixel in the calculation. The second method considers a 2×2 neighborhood, and it takes the weighted average of these four pixels to arrive at its final interpolated value. The latter method simply selects the value of the nearest point, and does not consider the values of other neighboring points. To find the most appropriate interpolation method, i.e. the method introducing the least error, several tests have been carried out. Otherwise standard parameters from Matlab were used, including anti-aliasing, which reduces artifacts in the downsampling procedure [9].

Evaluation of Different Transformation Types and Interpolation Methods

The five different transformation types and three different interpolation types, resulting in 15 combinations, should be evaluated the find the one with the lowest error. These combinations are applied to three different digital test images, as shown in Figure 4. Before applying these combinations, the digital image going to be registered, has been rotated and resized so that the most relevant problems during the scanning procedure is faced. In this process no scanning has been carried, only a "simulation" of the issues encountered from the scanning. Finally, the results have been compared with the original image in order to calculate the error introduced in the process.

(a) Image 1 (b) Image 2 (c) Image 3

Figure 4. *Images used for evaluating different transformation types and interpolation methods.*

The test images are selected to have a good combination of what would be expected for printed images. The first image (Figure 4(a)) is a typical scenery, the second image (Figure 4(b)) is an image with a fairly uniform background, and the third image (Figure 4(c)) is an image with fine details. Insisting on the uniformity and non-uniformity of the background is because edges have a great influence on the errors when comparing the original and the registered image.

Figures 5 and 6 show the Mean Square Error (MSE) between the original image and the registered image after applying the combinations of different transformation and interpolation types for Figure 4(a) and Figure 4(c). Results for Figure 4(b) are similar to the results for Figure 4(a), and is therefore not shown. It should be kept in mind that the MSE values in the figures are normalized with the highest value for the combinations. The results are also similar for the computed absolute difference values between the original and registered images.

As it can be seen, using a "similarity" transformation and a "bilinear" interpolation has the lowest MSE value. Inspection of the difference between the original image and the registered image show that the errors using this combination of transformation

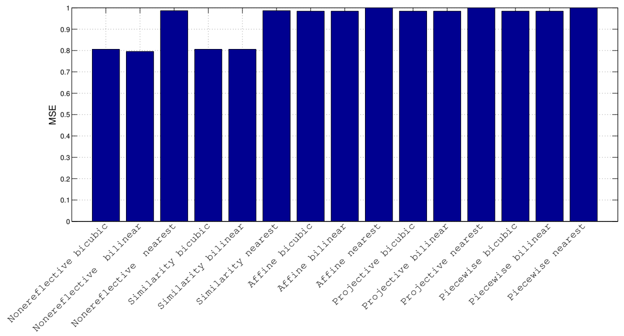

Figure 5. MSE between the original and registered image calculated for Figure 4(a). Values are normalized by the highest value.

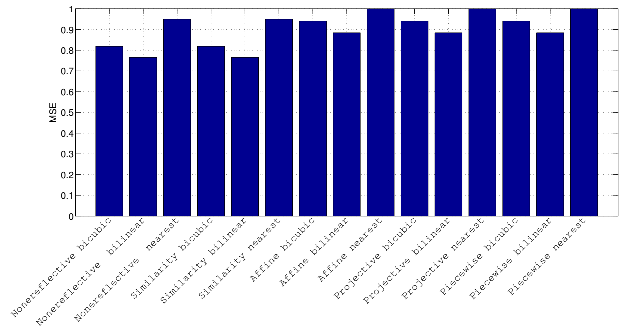

Figure 6. MSE between the original and registered image calculated for Figure 4(c). Values are normalized by the highest value.

and interpolation occur on the edges. In addition to our results, research carried out by Acharya and Tsai [10] indicates that bilinear interpolation has the least error when reducing the size of an image, supporting our findings.

Comparison Against Another Framework

The most important aspect to consider when selecting a framework for color prints is that the errors introduced are as small as possible. From the state of the art two different types of frameworks can be found, those who use local interest points, such as the one proposed by Eerola et al. [5], and those who use control points, such as the framework by Yanfang et al. [4] and the modified framework presented above. Both have disadvantages and advantages, for the local features no extra control points needs to be added in the process, but local interest points will not work well for uniform or near-uniform surfaces, since no points can be found in these areas. Using control points can be a disadvantage, since these need to be added prior to printing. However, uniform images can be correctly registered, including color patches. We compare these two different types of IQ frameworks, the one proposed by Eerola et al. and the modified version of the framework introduced above.

We use the images in Figure 4 to compare the frameworks. For the proposed modified framework we use bilinear interpolation and similarity as transformation, for the framework by Eerola et al. we use the same settings as in their paper. The best framework should have the least difference between the original image and the registered image. The results for the images are shown in Figure 7, and we can clearly see that the proposed framework based on control points introduces less error than the framework by Eerola et al. based on local features. The biggest difference is found in the image with uniform areas, which is a problem for frameworks based on local features since no registrations points can be found. In images with a lot of details the difference in error is almost insignificant, but the proposed framework performs slightly better (Image 3 in Figure 7).

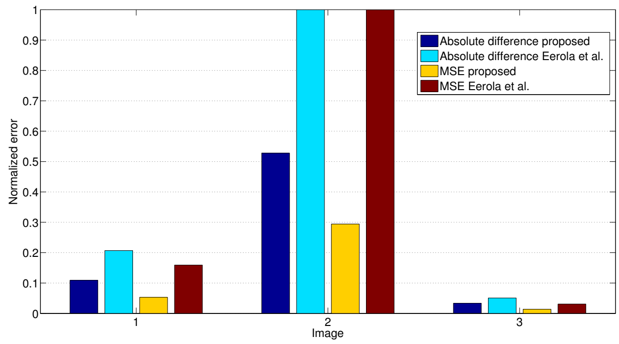

Figure 7. Error of the proposed framework using control points compared to the error from the framework from Eerola et al. [5] using local features. Errors are normalized by the highest value for both MSE and absolute difference.

Another important aspect for using the framework is the computational time, because of this we also investigated the time for the frameworks to register the three images used in the evaluation above. The computational time for the proposed framework is stable over the three images since it does not depend on the content of the image. The other framework, which is based on local features, has varying computational time since the content of the image affects the number of registration points. Comparing the time used by the two different framework, the proposed framework is more than 20 times faster than the framework by Eerola et al.. The time differences come mainly from the number of registration points. Furthermore, the complexity of a framework based on local features is higher than that of a framework based on control points.

Based on the results shown here, we will use the proposed modified framework based on control points to evaluate IQ of color prints.

Application of the Image Quality Framework

We have used the framework explained above to evaluate a set of IQ metrics on images from a color work flow.

Experimental Setup

15 images were obtained from Cardin [11] (Figure 8). These were processed with two different source profiles, the sRGB v2 perceptual transform and the sRGB v4 perceptual transform. These were further processed with four different softwares for obtaining the destination profile:

- Basic rendering (black point compensation and hue preserving minimum ΔE^*_{ab}).
- LOGO colorful from ProfileMaker Pro.
- LOGO Chroma plus from ProfileMaker Pro.
- Onyx from Onyx Mediacreator.

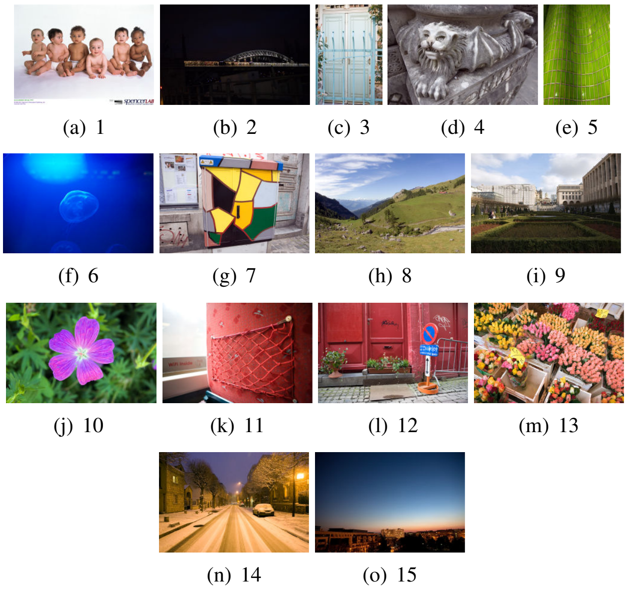

Figure 8. Images in the experiment.

For each image this results in eight different reproductions, as seen in Figure 9. These images were printed, with the Océ ColorWave 600 wide format CMYK printer on Océ Red Label paper, at a resolution of 150 pixels per inch and at a size resulting in a printed reproduction at approximately 8 by 10 inches.

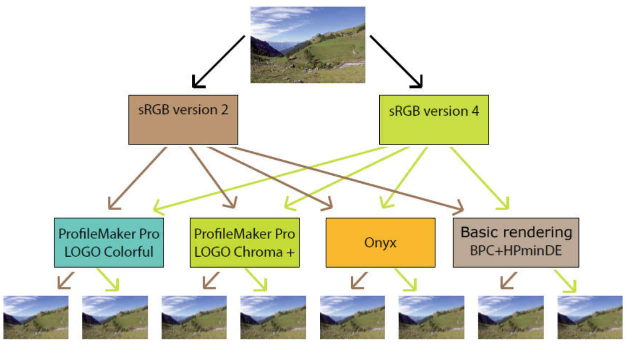

Figure 9. Overview of the work flows used in the experiment. The image is either processed using the sRGB v2 or sRGB v4. Furthermore, each of these is used as input to one of the four rendering methods. This finally results in eight different printed reproduction of the same image.

The printed images were evaluated by 30 observers in a category judgment experiment with three categories. The three categories where described as:

1. the most pleasant,
2. neither more nor less pleasant, and
3. less pleasant.

The observers were presented with a reference image on an EIZO ColorEdge CG221 display at a color temperature of 6500 Kelvins and luminance level of 80 cd/m^2, following the specifications of the sRGB. The image set was rendered for a sRGB display, and therefore a monitor capable of displaying the sRGB gamut was the most adapted reproduction device for this set of images. The printed images were presented randomly in a controlled viewing room at a color temperature of 5200 Kelvins, an illuminance level of 450 $\pm$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 [12] as closely as possible. More details on the experiment can be found in Cardin [11] or Bonnier et al. [13].

The 15 images printed with the eight different work flows were scanned with a Microtek ScanMaker 9800XL at 600dpi without any automatic corrections. The scanner was characterized using a Kodak IT8.7 target, since a printed test target was not available, and the profile was build using ProfileMaker 5.0.8. Analysis of the scans showed that the scanner exhibit a slight shearing, which occurs due to mechanical issues. Because of this a transformation which takes into account shearing must be applied. In scanners without shearing a simpler transformation types just incorporating translation, rotation, and scaling can be used. The difference between affine (correcting for shearing) and similarity (not correcting for shearing) is small, and does not affect the results significantly. In this experiment we adapt affine transformations as the transformation method and bilinear interpolation as the interpolation method.

Results

30 observers participated in the experiment, where they judged the images on a three step scale. The answers from the observers where processed with the Colour Engeneering Toolbox [14] to obtain z-scores. The results are shown in Figure 10. Onyx V4 has the highest z-score, but cannot be differentiated from Onyx V2. It is worth noticing the small difference between the highest and lowest z-score. This indicates a low visual difference between the different work flows, and that the task were difficult for the observers. We will focus on the objective evaluation of IQ, but an in depth analysis of the subjective evaluation can be found in Cardin [11].

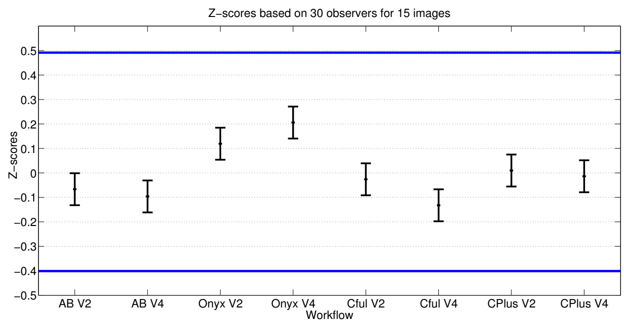

Figure 10. Z-scores based on 30 observers for 15 images. Solid horizontal lines indicate categories.

Image Quality Metrics

A set of IQ metrics were applied to evaluate the IQ of the printed images. Since no direct descreening is applied in the registration process IQ metrics not incorporating aspects of the HVS are inappropriate. We have selected IQ metrics both for overall IQ and for suitable quality attributes [15, 16]. For overall IQ the following metrics were used:

- Spatial CIELAB (S-CIELAB) [3] is often used as a reference metric, and has wide acceptance.
- S-CIELAB$_{Johnson}$ [17] is an improved version of the original S-CIELAB, in terms of the spatial filtering.

Since a color work flow was evaluated IQ metrics for assessing the color quality attribute is also relevant, and therefore the following metrics were used as well:

- Spatial-DEE (S-DEE) [18] extends the S-CIELAB$_{Johnson}$ with a more refined color difference calculation.
- Spatial Hue Angle MEtric (SHAME) [19] combines two state of the art metrics (the hue angle measure [20] and S-CIELAB$_{Johnson}$). We have included both SHAME-I and SHAME-II, which differ in terms of the spatial filtering.
- Adaptive Bilateral Filter (ABF) [21] uses bilateral filtering to simulate the HVS, before the color difference is calculated using ΔE^*_{ab}.

Further, IQ metrics taking into account structural information might be suitable, since these are potentially good at detecting artifacts.

- Structural SIMilarity (SSIM) [22], since this is commonly used and has received great publicity since it was proposed in 2004. This metric works on a local neighborhood, and is therefore considered as appropriate for our use.
- Cao et al. [23] proposed a metric designed to detect artifacts, which is based on the difference of saliency. This metric should be suitable since the dataset from Cardin reported to have different artifacts, such as loss of details and contouring [13]. The percentage of pixels with artifacts has been used as a measure of quality for this metric.

Evaluation of performance is done by calculating the correlation coefficient between the subjective score and the objective score. Three different kind of correlation are computed; the Pearson product-moment correlation coefficient, the Spearman's rank correlation coefficient, and the Kendall tau rank correlation coefficient [24]. The first assumes that the variables are ordinal, and finds the linear relationship between variables. The second, Spearman, is a non-parametric measure of correlation that uses the ranks as basis instead of the actual values. It describes the relationship between variables without making any assumptions about the frequency distribution of the variables. The third, Kendall, is a non-parametric test used to measure the degree of correspondence between two rankings, and assessing the significance of this.

Two types of evaluation is carried out, first the overall performance of the metrics over the entire image set, and then evaluation of the metrics for each of the 15 different images.

Overall Evaluation

We compare the overall score from the observers to the overall score by the metrics in order to evaluate the overall performance of the metrics. The most common method for this is by computing the Pearson correlation for all scores. The results from this show that all metrics have a very low Pearson correlation, approximately around zero. This indicates that the IQ metrics cannot predict perceived IQ. In addition, both the Spearman and

Metric	Correlation
S-CIELAB	0.34
S-CIELAB$_{Johnson}$	0.14
S-DEE	-0.27
SHAME-I	-0.29
SHAME-II	0.19
ABF	-0.09
SSIM	0.18
Cao et al.	-0.60

Table 2. Overall performance of the metrics using the method proposed by Pedersen and Hardeberg [25].

Kendall correlation coefficents are similar to the Pearson correlation.

The low correlation is because of scale differences between the images, and because of this problem a new method for evaluation of overall performance of metrics was proposed by Pedersen and Hardeberg [25]. In this method the scores for the metric are used to simulate an observer using the rank-order method, and results in a z-score plot as from the experiment (Figure 10). The results from the metric should be similar to the results from the observers if the metric exhibit a good performance.

Figure 11 shows the results for the S-CIELAB IQ metric with this method. A visual inspection between the results from the metric and the observers show differences, which is evidence that the metric does not predict perceived overall IQ. The Pearson correlation between these are only 0.34, indicating a low correlation. The other also have a low correlation, as seen in Table 2.

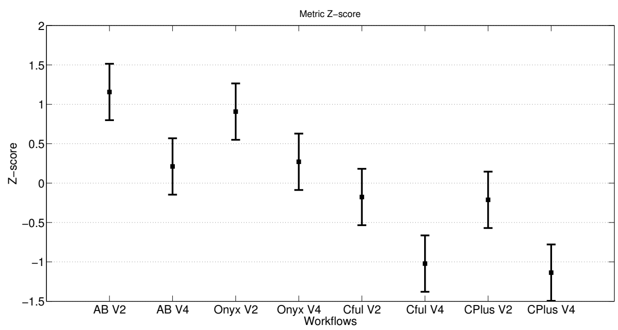

Figure 11. *S-CIELAB Z-score calculated with the method proposed by Pedersen and Hardeberg [25].*

The small visual quality differences between the images contribute to making this a difficult task for IQ metrics. For this experiment the IQ metrics cannot predict perceived overall IQ.

Image-wise Evaluation

We have also analyzed the performance of the IQ metrics for each image, since previous research by Hardeberg et al. [26] has shown that some IQ metrics can be linked with certain characteristics of the image. Image-wise evaluation will reveal what causes the low overall performance of the metrics. Figure 12 shows the Pearson correlation for the different IQ metrics for the 15 different images. We can notice a great variance in the results for the different images, and also some variance between the IQ metrics. The Spearman and Kendall correlation follow the same tendency as the Pearson correlation.

Figure 12 shows that metrics such as S-CIELAB, S-CIELAB$_{Johnson}$, SHAME-I, SHAME-II, and ABF perform similar in many of the images. In images 5, 7, 8, 11, and 12 they have a good correlation. S-DEE differs a bit from the previous metrics, most likely since it is based on another color difference formula. SSIM is different from the other metrics in terms of performance, and it has generally low correlation, except in images 8, 11, and 12. It should noted that SSIM is a gray scale metric, and does not take into color information, which can explain why it is worse than the other metrics. The metric by Cao et al. performs opposite of the other metrics, this is not surprising, since it is based on artifact detection. We can see a very high correlation in image 3, where the observers mostly judged the images based on detail visibility.

In the images where there is a large difference from the original, but the difference is increasing the quality, the IQ metrics do not perform well. In the second image some of the reproductions have lost shadow details, but they have a small difference from the original. In this image observers preferred reproductions where details were preserved at the cost of larger color differences. This is also the reason why the metric from Cao et al. is very similar to the observers. The problem of when a difference contributes to increasing IQ is a difficult issue to handle for the IQ metrics, and the IQ metrics are still not good enough to predict this. However, in the images where a small difference from the original is preferred, such as image 5, 7, 8, 11, and 12, the IQ metrics perform reasonably well. In these cases IQ metrics working solely on the difference from the original are giving good results. This indicates that the metrics could be chosen based on the characteristics of the image.

Conclusion

We have proposed a simple framework for using IQ metrics on printed images. It is a modified version of one of the proposed frameworks in the literature. The modified framework, which is based on control points, has lower complexity and is faster than a state of the art framework, which is based on local features. The experimental results also show that the proposed framework produces fewer errors to the registered image.

We have used the modified framework to evaluate a selection of IQ metrics on a set of images from different color work flows. Results indicate that the IQ metrics cannot predict perceived overall IQ, however, the results seem to be both metric and image dependent.

References

[1] 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), Jun 2009. ISSN: 1890-520X.

[2] X. Zhang, D.A. Silverstein, J.E. Farrell, and B.A. Wandell. Color image quality metric S-CIELAB and its application on halftone texture visibility. In *COMPCON97 Digest of Papers*, pages 44–48, Washington, DC, USA, 1997. IEEE Computer Society.

[3] X. Zhang and B.A. Wandell. A spatial extension of CIELAB for digital color image reproduction. In *Soc. Inform. Display 96 Digest*, pages 731–734, San Diego, 1996.

[4] X. Yanfang, W. Yu, and Z. Ming. Color reproduction quality metric on printing images based on the s-cielab model. In *2008 International Conference on Computer Science and Software Engineering*, pages 294–297, 2008.

[5] T. Eerola, J-K. Kamarainen, L. Lensu, and H. Kalviainen. Framework for applying full reference digital image quality measures to printed images. In *Scandinavian Conference on Image Analysis*, Oslo, Norway, June 2009.

[6] A. Eskicioglu. Quality measurement for monochrome compressed images in the past 25 years. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pages 1907–1910, 2000.

[7] W. Lim and S. Mani. Application of digital imaging to measure print quality. In *International Conference on Digital Printing Technologies*, pages 611–614, 1998.

[8] A. Goshtasby. Piecewise linear mapping functions for image registration. *Pattern Recognition*, 19:459–466, 1986.

[9] J. A. Parker, R.V. Kenyon, and D. E. Troxel. Comparison of interpolating methods for image resampling. *IEEE Transaction on medical imaging*, MI-2(1):31–39, March 1983.

[10] T. Acharya and P-S. Tsai. Computational foundations of image interpolation algorithms. *Ubiquity*, 8:1–17, October 2007. doi: http://doi.acm.org/10.1145/1317487.1317488.

[11] N. Cardin. L'utilisation du perceptual reference medium gamut dans la gestion des coleours améliore-t-elle la qualité d'images produites par impression jet d'encre? Master's thesis, École Nationale Supérieure Louis-Lumière Promotion Photographie, 2009.

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

[13] N. Bonnier, P. Green, and A. Sarlat. Evaluating the use of the perceptual reference medium gamut in ICC printing works. In *Seventeenth Color Imaging Conference*, pages 227–232, Albuquerque, NM, Nov. 2009. IS&T and SID.

[14] P. Green and L. MacDonald, editors. *Colour Engineering: Achieving Device Independent Colour*. John Wiley & Sons, 2002.

[15] 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 – 011016–13, Jan 2010.

[16] 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, New Mexico, USA, Nov 2009.

[17] G. M. Johnson and M. D. Fairchild. Darwinism of color image difference models. In *The 9th Color Imaging Conference: Color Science and Engineering: Systems, Technologies, Applications*, pages 108–112, 2001.

[18] G. Simone, C. Oleari, and I. Farup. Performance of the euclidean color-difference formula in log-compressed osa-ucs space applied to modified-image-difference metrics. In *11th Congress of the International Colour Association (AIC)*, Sydney, Australia, Oct 2009.

[19] M. Pedersen and J. Y. Hardeberg. A new spatial hue angle metric for perceptual image difference. In *Computational Color Imaging*, volume 5646 of *Lecture Notes in Computer Science*, pages 81–90, Saint Etienne, France, Mar 2009. Springer Berlin / Heidelberg. ISBN: 978-3-642-03264-6.

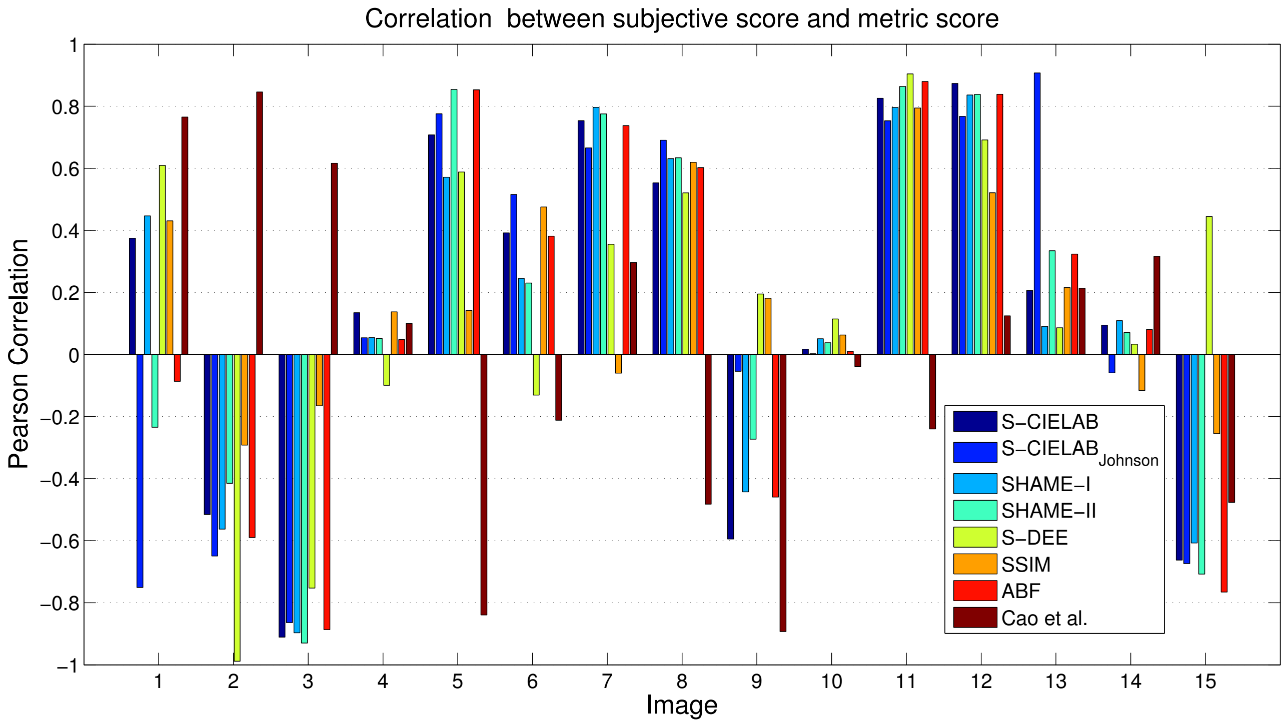

Figure 12. *Pearson correlation image-wise. For some images some IQ metrics have a high correlation, but the results vary over the 15 images in the dataset. The Spearman and Kendall correlation follow the same tendency as the Pearson correlation.*

[20] G. Hong and M.R. Luo. Perceptually based colour difference for complex images. In R. Chung and A. Rodrigues, editors, *9th Congress of the International Colour Association*, volume 4421 of *Proceedings of SPIE*, pages 618–621, 2002.

[21] Z. Wang and J. Y. Hardeberg. An adaptive bilateral filter for predicting color image difference. In *17th Color Imaging Conference*, pages 27–31, Albuquerque, NM, USA, Nov 2009.

[22] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004.

[23] G. Cao, M. Pedersen, and Z. Baranczuk. Saliency models as gamut-mapping artifact detectors. In *CGIV*, Joensuu, Finland, Jun. 2010.

[24] M. G. Kendall, A. Stuart, and J. K. Ord. *Kendall's Advanced Theory of Statistics: Classical inference and relationship*, volume 2. A Hodder Arnold Publication, 5 edition, 1991.

[25] M. Pedersen and J. Y. Hardeberg. Rank order and image difference metrics. In *CGIV 2008 Fourth European Conference on Color in Graphics, Imaging and Vision*, pages 120–125, Terrassa, Spain, Jun 2008. IS&T.

[26] J.Y. Hardeberg, E. Bando, and M. Pedersen. Evaluating colour image difference metrics for gamut-mapped images. *Coloration Technology*, 124(4):243–253, Aug 2008.

Acknowledgments

We would like to thank Jon Yngve Hardeberg for the enlightening discussions.

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.

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 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.*

Seyed Ali Amirshahi *received his BsC in Electrical Engineering in 2008 from Amirkabir University of Technology (Tehran Polytechnic), Iran and he is currently pursuing a master in Color in Informatics and MEdia Technology (CIMET).*