Image Quality Measures for Evaluating Gamut Mapping

Zofia Barańczuk and Peter Zolliker, Laboratory for Media Technology; Swiss Federal Laboratory for Materials Testing and Research (EMPA), Dübendorf, Switzerland
Joachim Giesen, Friedrich-Schiller-Universität Jena, Germany

Abstract

In this paper we compare different image quality measures for the gamut mapping problem, and validate them using psycho-visual data from four recent gamut mapping studies. The psycho-visual data are choice data of the form: given an original image and two images obtained by applying different gamut mapping algorithms, an observer chooses the one that reproduces the original image better in his/her opinion. The scoring function used to validate the quality measures is the hit rate, i.e., the percentage of correct choice predictions on data from the psycho-visual tests. We also propose a new image quality measure based on the difference in color and local contrast. This measure compares well to the measures from the literature on our psycho-visual data. Some of these measures predict the observer's preferences equally well as scaling methods like Thurstone's method or conjoint analysis that are used to evaluate the psycho-visual tests. This is remarkable in the sense that the scaling methods are based on the experimental data, whereas the quality measures are independent of this data.

Introduction

Gamut mapping describes how a color image is rendered on a device with limited color reproduction capabilities. This classical problem is still an area of active research – Morovic gives a good recent overview [1]. An important step in improving gamut mapping algorithm is an accurate evaluation of its psycho-visual performance. This is traditionally achieved using psycho-visual tests, where observers have to decide which of the mapped images are the better representation of the original. The data gathered in such a test are typically evaluated using Thurstone's Law of Comparative Judgement [2]. An alternative approach that we want to evaluate here is to use an image quality measure (independent of observer feedback) to measure the difference of a mapped image to the original. An overview of the state of the art in image quality research can be found for example in [3] or [4]. Image quality measures are successfully used in many imaging applications, such as modeling image distortions, especially in data compression [5]. The advantage of using "good" image quality measures to evaluate gamut mapping algorithms is that they can be used to automatically predict the perceived quality of a mapped image without the need for a new psycho-visual study. Psycho-visual tests generally give reliable results for tested settings but the tests are time consuming. Furthermore, an extrapolation to changed settings and new images is problematic. Computing an image quality measure on the other hand provides results immediately. The challenge is to find a measure that correlates well with observers' preferences. It has to represent the response of the human visual system as a mathematical function.

For gamut mapping the main image quality factors are preservation of lightness/color and preservation of spatial details. Artifacts introduced by the mapping algorithms may also be a factor which, however, will be neglected in the present study. Some factors encountered in other image quality applications such as noise or compression artifacts are of minor importance for gamut mapping.

The main topic of this paper is a quantitative comparison of the performance of image quality measures with data driven quality measures from psycho-visual tests. The performance of the measures is assessed as the percentage of correctly predicted observer choices on data from a psycho-visual test. The data used to compute these percentages were neither used for data evaluation nor for the optimization of the corresponding measures (compare Section "Validating the quality measures"). Correlations of psycho-visual gamut mapping evaluation and image quality measures have been published before in [6], where only a general ranking of gamut mapping algorithms has been discussed. Our focus here is on predicting observers' choices in individual comparisons between mapped images.

The remainder of this paper is organized as follows: In the next two sections we describe the image quality measures considered in this paper. Then Thurstone's method and an extension to conjoint analysis are briefly described as methods for evaluating psycho-visual test data. In the subsequent section we describe how to validate the different image quality measures for gamut mapping. The data sets which we used for validation are described in a section on its own. Finally, we discuss the experimental validation results on data sets and conclude the paper.

Image quality measures

In this section we review the image quality measures that we have compared. We always compare two images X and Y with $n \times m$ pixels. At the pixels $x_{ij} \in X$ and $y_{ij} \in Y$, respectively, we consider color coordinates. Mostly we are using the lightness coordinate L in CIELAB color space. If not stated otherwise we do not distinguish in our notation between a pixel and the color coordinate considered at this pixel.

Structural Similarity Index (SSIM)

The *Structural Similarity Index* was introduced by Wang et. al. [7] and is defined on quadratic image patches of size $k \times k$ at the same location within image X and Y. We computed SSIM for the L coordinate in CIELAB color space. Let $P_X \subset X$ be such a patch and P_Y the corresponding patch for Y. We compute the following quantities for the patches:

$$\bar{P}_X = \frac{1}{k^2} \sum_{x \in P_X} x, \quad \bar{P}_Y = \frac{1}{k^2} \sum_{y \in P_Y} y,$$

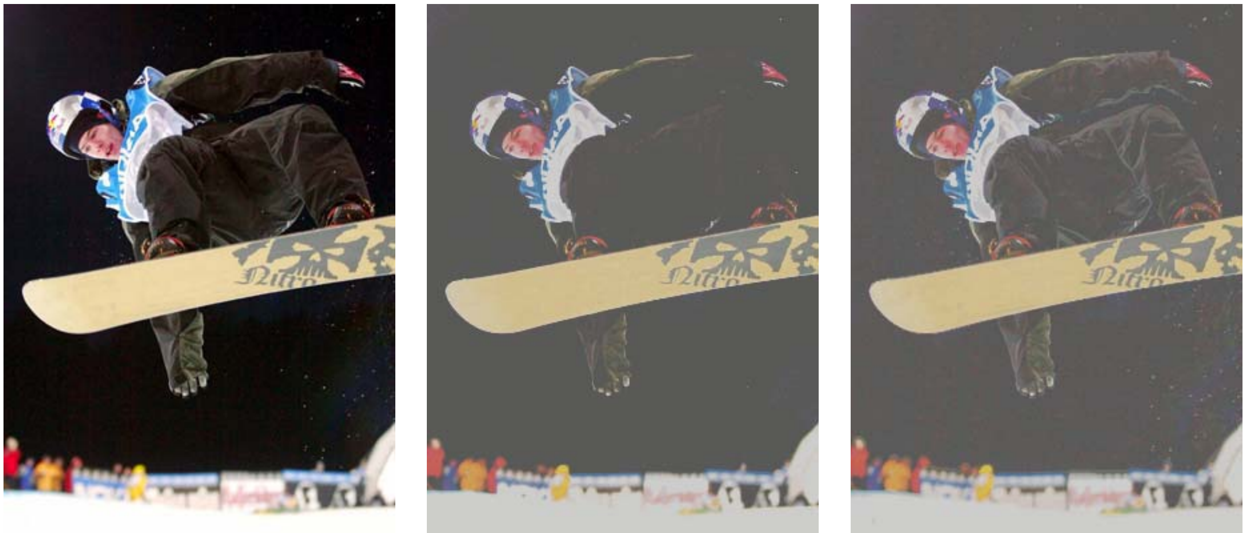

Figure 1: The original image (on the left) and two gamut mapped images (in the middle and on the right). For the image in middle we have $Q_{\Delta E} = 24.65$ and $Q_{\Delta LC} = 0.341$ using HPminDE without detail enhancement, and on the right we have $Q_{\Delta E} = 27.00$ and $Q_{\Delta LC} = 0.318$ using HPminDE with details enhancement. Note for the image in the middle $Q_{\Delta E}$ is smaller than for the image on the right, but the middle image has lost a lot of details and has the larger perceptual distance from the original (left image).

$$\begin{aligned}
{\sigma_{P_X}}^2 &= \frac{1}{k^2-1}\sum_{x\in P_X}(x-\bar{P}_X)^2, \\
{\sigma_{P_Y}}^2 &= \frac{1}{k^2-1}\sum_{y\in P_Y}(y-\bar{P}_Y)^2, \text{ and} \\
\sigma_{P_X P_Y} &= \frac{1}{k^2-1}\sum_{i=1}^{k^2}(x_i-\bar{P}_X)(y_i-\bar{P}_Y)
\end{aligned}$$

The Structural Similarity Index is then defined as

$$\mathrm{SSIM}(P_X,P_Y) = \frac{(2\bar{P}_X\bar{P}_Y+c_1)(2\sigma_{P_X P_Y}+c_2)}{\left(\bar{P}_X^2+\bar{P}_Y^2+c_1\right)\left(\sigma_{P_X}^2+\sigma_{P_Y}^2+c_2\right)},$$

with two constants c_1 and c_2. As proposed by Wang et. al. [5] we used $c_1 = 1$ and $c_2 = 9$ for these constants and $k = 8$ for the patch size.

From the Structural Similarity Index the image quality measure $Q_{\mathrm{SSIM}}(X,Y)$ can be defined as the Structural Similarity Index SSIM averaged over all possible $k\times k$ patches in the images X and Y. The resulting measure is in the range $[-1,1]$, and the higher the Q_{SSIM} value, the more similar are the compared images.

Laplacian mean square error (LMSE)

Like the Structural Similarity Index the Laplacian Mean Square Error (compare [8]) is a local measure for the difference in two images. We compute the following quantities at each pixel (more exactly at L coordinate in CIELAB color space of each pixel, with indices $2 \le i \le n-1$ and $2 \le j \le m-1$) of X and Y, respectively:

$$\begin{aligned}
L(x_{ij}) &= x_{(i+1)j}+x_{(i-1)j}+x_{i(j+1)}+x_{i(j-1)}-4x_{ij} \\
&\text{and} \\
L(y_{ij}) &= y_{(i+1)j}+y_{(i-1)j}+y_{i(j+1)}+y_{i(j-1)}-4y_{ij}
\end{aligned}$$

The image quality measure Q_{LMSE} is then defined as

$$Q_{\mathrm{LMSE}}(X,Y) = \frac{1}{(n-2)(m-2)}\sum_{i=2}^{n-1}\sum_{j=2}^{m-1}\left(L(x_{ij})-L(y_{ij})\right)^2.$$

Mean square error (MSE)

We also consider the mean square error which is just the squared pointwise difference between the images X and Y. The corresponding image quality measure Q_{MSE} is defined as

$$Q_{\mathrm{MSE}}(X,Y) = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}(x_{ij}-y_{ij})^2,$$

where x_{ij} and y_{ij} are L coordinates in the CIELAB color space for the points in images X and Y respectively.

Discrete wavelet transform (DWT)

The discrete wavelet transform image quality measure has been defined in [9]. Images X and Y are compared as follows: a discrete wavelet transform is applied to the luminance layer of image X and Y, respectively. Let M_X^f be the magnitudes of the discrete wavelet transform coefficients obtained for X and frequency band f, and let M_Y^f be the corresponding magnitudes for image Y. From M_X^f and M_Y^f the absolute values of differences

$$d_i^f(X,Y) = \left|M_{X\,i}^f - M_{Y\,i}^f\right|, \quad i = 1,\ldots,\left|M_X^f\right| = \left|M_Y^f\right|.$$

are computed for each frequency band. Let $\sigma_f(X,Y)$ be the standard deviation of the differences $d_i^f(X,Y)$ for frequency band f. Now, the $Q_{\mathrm{DWT}}(X,Y)$ image quality measure is defined as the mean of the $\sigma_f(X,Y)$ for all the frequency bands.

In our implementation we use Daubechies' filter [10] to compute the discrete wavelet transform image quality measure Q_{DWT}.

A new quality measure

The quality measure that we are going to describe in this section is based on the observation that important factors determining the quality of gamut mapping algorithms are color preservation and contrast (detail) preservation. We estimate the degree of color preservation by using the CIELAB ΔE distance measure [11]. The example in Figure 1 demonstrates that $Q_{\Delta E}$, i.e., the quality measure derived from ΔE, alone is not an accurate quality measure since it neglects the preservation of details. To account also for detail preservation we introduce a contrast preserving measure that we call $Q_{\Delta LC}$. Our quality measure is then a linear combination of $Q_{\Delta E}$ and $Q_{\Delta LC}$. Below we describe the two measures $Q_{\Delta E}$ and $Q_{\Delta LC}$ in more detail.

The measure $Q_{\Delta E}$

ΔE is defined as the Euclidean distance in CIELAB color space between corresponding pixels in two images X and Y of size $n \times m$. That is, locally at pixel $x \in X$ and the corresponding pixel $y \in Y$ the ΔE distance is defined as:

$$\Delta E(x,y) = \sqrt{((L_x - L_y)^2 + (a_x - a_y)^2 + (b_x - b_y)^2)}$$

As our image quality measure $Q_{\Delta E}$ we take the average ΔE over the pixels of the two images, i.e.,

$$Q_{\Delta E}(X,Y) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \Delta E(x_{ij}, y_{ij}).$$

ΔE is a popular image quality metrics since it is easy to compute and has a natural interpretation, though in principle it could be replaced by any more sophisticated color distance measure such as CIECAM02 [12], or ΔE_{94} [13].

The measure $Q_{\Delta LC}$

The image quality measure $Q_{\Delta LC}$ is based on a local contrast measure. We chose the Michelson contrast, see [14], as our measure of local contrast. We compute the Michelson contrast on a $k \times k$ patch $P_X \subset X$ of the image X as follows (we were using $k = 5, 17$ and 33 in our experiments):

$$LC(P_X) = \frac{x_{max} - x_{min}}{x_{max} + x_{min}},$$

where x is an luminosity coordinate in XYZ color space (at pixel $x \in P_X$), and x_{max} and x_{min} are the highest value and the lowest value, respectively, of this intensity on the patch P_X. Analogously, we can compute the value $LC(P_Y)$ for the corresponding patch P_Y in image Y, and define

$$\Delta LC(P_X, P_Y) = |LC(P_X) - LC(P_Y)|.$$

The image quality measure $Q_{\Delta LC}(X,Y)$ is then finally defined as the measure ΔLC averaged over all possible $k \times k$ patches in images X and Y.

Thurstone's method and conjoint analysis

Traditionally image quality in gamut mapping is evaluated using Thurstone's Law of Comparative Judgement, which can be used to analyze paired comparison data [2]. Applying Thurstone's law allows to derive a value for each tested gamut mapping algorithm. Thurstone's method has been extended to a *conjoint analysis* of parameterized gamut mapping algorithms [15, 16]. In a conjoint analysis we still have paired comparison data, but instead of assigning a single value to a mapping algorithm we now assign it to all the parameters employed in the algorithms and sum up those scale values.

We use both methods to obtain image quality measures for images mapped with a gamut mapping algorithm on a image test set. These measures serve as a reference for the image quality measures discussed before, which do not need observer feedback in contrast to Thurstone's method or conjoint analysis.

To improve the consistency of results obtained by Thurstone's method or conjoint analysis, one can individualize the evaluation for each image. Individualization linearly combines Thurstone's scale values for the entire data set with scale values obtained separately for each image. The linear combination of those two scale values is then optimized to hold out data using cross validation. It turns out that scale values computed individually for images can be good, but typically they can be improved by shrinking them towards the scale values computed on the whole population of images (simply because in most cases there are not enough paired comparisons per image available). In a nutshell, the idea behind the individualization approach is providing a fallback when only few paired comparisons are available for an image. We will refer to these method as individualized Thurstone's method and individualized conjoint analysis, respectively.

Validating the quality measures

We need to validate the suitability of an image quality measure for gamut mapping. Our validation procedure estimates how well the quality measures align with observer ratings which we obtained in psycho-visual tests. As we have mentioned before, the data we elicited in the psycho-visual test are of the form: given an original image and two images obtained by applying different gamut mapping algorithms, a user chooses the one that reproduces the original image better in his/her opinion. We validate an image quality measure now by the percentage of correctly predicted observer choices. This validation measure is also known as hit rate. When computing hit rates for Thurstone's method or conjoint analysis we need to be careful that we do not validate the methods on the same data which we used to teach the model (remember that Thurstone's method and conjoint analysis are, in contrast to the other image quality measures, observer data driven). To circumvent this problem we use cross validation, i.e., we use part of the data to teach the models and the remaining part to validate the models. In the following we give more detail on how to compute hit rates and on how we employed cross validation.

Hit rate

For each paired comparison in a psycho-visual test we know the choice of the observer. In some tests we allowed ties, i.e., neither of the two options was preferred. We omit such ties from further analysis. Let C be the set of non-tied observer choices. For an image quality measure we always predict the choice with the higher value for this measure on the elements in C. Let $S \subseteq C$ be the subset of correctly predicted choices. Then the hit rate is defined as

$$HR = \frac{|S|}{|C|},$$

where $|S|$ and $|C|$ are numbers of elements in sets S and C, respectively.

Cross Validation

For Thurstone's method and for conjoint analysis we use cross validation. That means, that the set C of non-tied observer choices is partitioned randomly into ten subsets of equal size. Out of the ten subsets, one is retained for validating the model, and the remaining nine subsets are used as training data. This is repeated ten times, where each subset is used as the validation set once. The mean hit rate on the ten validation data sets is used as validation quality measure. In Figure 3 we refer to this method as *Thur gen*.

For the individualized Thurstone's method, we carried out a double cross validation. For double cross validation we use eight of the ten subsets as training set, one as optimization set, and the remaining one for validation. We compute general and individual scale values by Thurstone's method on the training set. Then we optimize the weights for the linear combination of the population and individualized scale values using an optimizing set. Finally, we use the hit rate on the validation set. We repeat this process 250 times and use the mean of the mean hit rates as validation quality measure. We refer to this method in Figure 3 as *Thur spec*.

Data sets

The different image quality measures were validated on image data of four recent gamut mapping studies. All tests used paired comparison, where two mapped images were compared to an original image. Three of the tests were carried out in a laboratory environment following the CIE guidelines [17] with ISONewspaper gamut as the target gamut. The fourth study was on a parameterized gamut mapping algorithm [16, 18] and the major part of it has been carried out over the internet. Detailed data on the size of these studies are summarized in Table 1. In the following we summarize the main ideas of the four studies.

Study 1: Basic Study (BS)

This study is a traditional benchmark study comparing some newer image dependent gamut mapping algorithms to known reference algorithms. In addition to the reference algorithms HPminDE, SGCK [17], the following algorithms using image gamut or spatial gamut mapping have been considered: the algorithm NOptStar that is using the image gamut as described in [19], the Kolas algorithm [20], the Zolliker algorithm [21] applied to the SGCK and NOptStar algorithms, and the Caluori algorithm [22]. For this study 97 images were used, each mapped with seven algorithms. Each possible comparison was made at least once. We will refer to this study as *Basic study* or simply *BS*.

Study 2: Image Gamut (IG)

The topic of this study was the use of image gamut descriptions for gamut mapping [19]. We considered algorithms using a linear and sigmoidal mapping, each of them had three possible source gamuts, namely device gamut (sRGB) and two types of image gamut description. The six possible combinations were compared to HPminDE and SGCK, resulting all together in eight algorithms. 75 images were used. Each possible comparison was made approximately twice. We will refer to this study as *Image Gamut study* or simply *IG*.

Study 3: Local Contrast (LC)

In this study the influence of detail enhancement applied to a set of gamut mapping algorithms was investigated [21]. The study comprised the HPminDE, SGCK, SGDA [23] algorithms and a linear compression algorithm. All algorithms were compared with and without detail enhancement. 77 images were used, and 5376 comparisons have been performed. Each possible comparison was made approximately 2.5 times. We will refer to this study as *Local Contrast study* or simply *LC*.

Study 4: Parameterized Gamut Mapping (PGM)

In this study ([15, 16]) a master algorithm with a set of parameters was studied. The parameters include compression type, detail enhancement, color space, gamut size as well as color, lightness and hue shifts. 97 images were used. Due to parameterization the number of possible algorithms per image was as high as 1536, and 5058 comparisons have been elicited. For the evaluation conjoint analysis was used. We will refer to this study as *Parameterized Gamut Mapping study* or simply *PGM*.

Study	Images	Comparisons	Algorithms
BS	97	2086	7
LC	77	5376	8
IG	75	4360	8
PGM	97	5058	1536

Table 1: Number of images, comparisons and algorithms in the four considered studies.

Experimental results

The main results of validating the different image quality measures on the four psycho-visual tests are summarized in Figure 3. Additionally, in Figure 2 we present the hit rates for individualized Thurstone's method or conjoint analysis. As expected, hit rates for training sets are higher than for test sets. Individualization improves training set hit rates, however this does not always translate to improvement over the test set.

Overall, the Structural Similarity Index measure (SSIM) proved to be the best performing image quality measure. On the BS, LC and IG studies it scores higher than the competing measures, and on the PGM data it is close behind the best of the other methods (our new measure – a combination of $Q_{\Delta E}$ and $Q_{\Delta LC}$[with $k = 5$]).

Interestingly, the results obtained with Thurstone's method are not significantly better than those resulting from image quality measures. In particular, results for Thurstone's method without individualization are comparable to results obtained for SSIM, i.e., the hit rates differ by only up to two percents. On BS data, the results obtained with SSIM are even better than those coming from Thurstone's method.

On the studies LC and IG the individualized Thurstone's method gives better results than the competing measures, but not on the remaining two studies BS and PGM. A probable reason for this behavior is the size of the studies. The performance of the individualized Thurstone's method improves with increasing number of comparisons, and image quality measures can compete with it as long as the number of comparisons is relatively small.

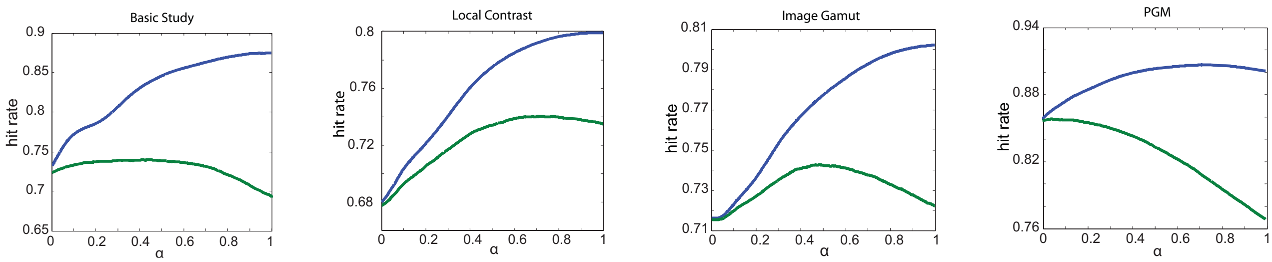

Figure 2: Hit rates using Thurstone's method or conjoint analysis for: (a) Basic Study, (b) Local Contrast Study, (c) Image Gamut Study, and (d) Parameterized Gamut Mapping Study. The blue (higher) line shows the hit rate on the training set, the green (lower) line shows the hit rate on the test set. Scale values (sv) are computed as a convex combination of scale values for the whole population of images (sv_{group}) and scale values for individual images (sv_{ind}), i.e., $sv = \alpha \cdot sv_{ind} + (1-\alpha) \cdot sv_{group}$ with $\alpha > 0$.

For the LC and IG studies the same comparison was tested approximately 2 and 2.5 times respectively more than for the BS study, thus giving more information for the individual images. Hence the results for individual images obtained from individualized Thurstone's method are more exact on these studies than on the other studies. However, individualized Thurstone's method is extremely inefficient in a sense that it requires a lot of time and observers for testing. On the BS study, where the amount of comparisons per image is the lowest, the individualized version of Thurstone's method is hardly better than the ordinary Thurstone's method.

On the PGM study all investigated image quality measures give better results than the individualized conjoint analysis, i.e., for this study it is more difficult to predict choices using conjoint analysis and easier using image quality measures. On this study all computed hit rates are relatively high compared to the other studies. This is due to the relatively large differences between the images used for that study, i.e., the decisions in the paired comparisons were easier. At the same time, this study required conjoint analysis to compute scale values for all algorithms. Therefore more comparisons were needed to obtain statistically significant results and apparently the number of comparisons per image was not high enough to yield a higher hit rate than for the global evaluation.

We considered two pointwise image quality measures, namely $Q_{\Delta E}$ and the mean square error Q_{MSE}. On all studies except PGM, these measures scored lower than their competitors, often showing hit rates close to random choice, i.e., 50%. This is because all gamut mapping algorithms tested in these studies already optimize color preservation in one way or the other. Hence, observers' choices are more affected by detail preservation. In particular, clipping algorithms, for example the HPminDE algorithm, are optimizing the transformed image against the pointwise distance measure, but ignores the detail preservation issue.

As noted above, the pointwise measures perform better on the PGM study. The reason is that many paired comparisons included a choice among different destination gamut sizes. Large differences in the size of the destination gamut have a large impact on the perceived image quality and at the same time are strongly correlated with $Q_{\Delta E}$. The differences in $Q_{\Delta E}$ of mapped images are higher, and they are well correlated with the perceived quality of the image. A suitable combination of $Q_{\Delta E}$ and $Q_{\Delta LC}$ scored highest on this study, even higher than the SSIM measure.

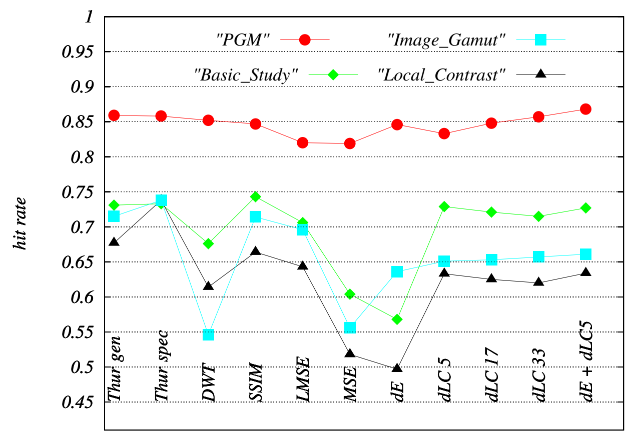

Figure 3: Hit rates obtained by different methods for four studies.

On each study the results obtained for the best measures are relatively similar, i.e., differences are within 5% in hit rate. The theoretical hit rate limit of 1.0 cannot be reached, because observers typically differ in their choices and even the decisions of a single observer may be inconsistent, i.e., the same person, under the same conditions can make a different choice on the same images in repeated paired comparisons.

Conclusions

We showed that image quality measures can be a useful and efficient method to gauge the quality of mapped images in gamut mapping. Overall, the best performing measure was the Structural Similarity Index (SSIM): it predicts choices of respondents similarly successfully as Thurstone's method. Better predictions can be achieved by computing individualized Thurstone's scale values, but only if enough test data is available. Also, simple combinations of measures of color distance ($Q_{\Delta E}$) and detail preservation ($Q_{\Delta LC}$) is very promising. We can expect a further improvement when we combine it with a measure for gamut mapping artifacts.

We can conclude that image quality measures like SSIM can be used for predicting choices in psycho-visual tests concerning the evaluation of gamut mapping algorithms. Evaluating gamut mapping algorithms automatically by using image quality measures can be an attractive alternative to psycho-visual test whenever the latter are too expensive or difficult to carry out.

Acknowledgments

This research was supported by the Hasler Stiftung Proj. # 2200.

References

[1] J. Morovic. *Colour Gamut Mapping*. WileyBlackwell, ISBN 0470030321, 2008.

[2] Peter G. Engeldrum. *Psychometric Scaling, A Toolkit for Imaging Systems Development*. Imcotek Press, Winchester MA, USA, 2000.

[3] B. Keelan. *Handbook of Image Quality*. Marcel Dekker, Inc., 2002.

[4] J. Dijk. *In search of an Objective Measure for the Perceptual Quality of Printed Images*. PhD thesis, Delft University of Technology, 2004.

[5] 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, April 2004.

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

[7] Z. Wang and A.C. Bovik. A Universal Image Quality Index. *IEEE Signal Processing Letters*, 9(3):81–84, March 2002.

[8] A.M. Eskicioglu and P.S. Fisher. Image Quality Measures and Their Performance. *IEEE Transactions on Communications*, 43(12):2959–2965, December 1995.

[9] D. Gayle, H. Mahlab, Y. Ucar, and A.M. Eskicioglu. A full-reference color image quality measure in the DWT domain. In *13th European Signal Processing Conference*, September 2005.

[10] G. Strang and T. Nguyen. *Wavelets and Filter Banks*. Wellesey, Wellesey-Cambridge Press, 1997.

[11] Central Bureau of the CIE, Vienna. *CIE Publication 15.2: Colorimetry*, 1986.

[12] N. Moroney, M.D. Fairchild, R.W.G. Hunt, C. Li, M.R. Luo, and T. Newman. The CIECAM02 Color Appearance Model. In *10th Color Imaging Conference*, pages 23–27. IS&T/SID, Nov 2002.

[13] R. McDonald and K.J. Smith. CIE94 - a new color-difference formula. *Journal of the Society of Dyers & Colourist*, 111(12):376–379, 1995.

[14] A. A. Michelson. *Studies in Optics*. U. Chicago Press, 1927.

[15] Z. Baranczuk, P. Zolliker, I. Sprow, and J. Giesen. Conjoint analysis of parametrized gamut mapping algorithms. In *16th Color Imaging Conference*, pages 38–43. IS&T/SID, Nov 2008.

[16] P. Zolliker, Z. Baranczuk, I. Sprow, and J. Giesen. Conjoint analysis used for the evaluation of parameterized gamut mapping algorithms. *accepted to IEEE Transactions in Image Proceeding*, 2009.

[17] Central Bureau of the CIE, Vienna. *CIE Publication 156: Guidelines for the Evaluation of Gamut Mapping Algorithms*, 2004.

[18] Iris Sprow, Zofia Baranczuk, Tobias Stamm, and Peter Zolliker. Web-based psychometric evaluation of image quality. In Susan P. Farnand and Frans Gaykema, editors, *Image Quality and System Performance VI*, volume 7242, page 72420A. SPIE, 2009.

[19] J. Giesen, E. Schuberth, K. Simon, and P. Zolliker. Image-dependent gamut mapping as optimization problem. *IEEE Transactions on Image Processing*, 16(10):2401–2410, October 2007.

[20] O. Kolas and I. Farup. Efficient hue-preserving and edge-preserving spatial color gamut mapping. In *Proc. 15th Color Imaging Conference Final Program and Proceedings*, pages 207–209, November 2007.

[21] P. Zolliker and K. Simon. Retaining local image information in gamut mapping algorithms. *IEEE Transactions on Image Processing*, 16(3):664–672, March 2007.

[22] U. Caluori and K. Simon. An RGB Color Management Concept based on an Improved Gamut Mapping Algorithm. In *Color Imaging XIV: Processing, Hardcopy and Applications*, volume 7241, page 724118A. SPIE, 2009.

[23] P. Zolliker and K. Simon. Continuity of gamut mapping algorithms. *Journal of Electronic Imaging*, 15(1):13004, March 2006.

Author Biography

Zofia Barańczuk received her bachelor's degree in computer science and master's degree in mathematics from the Warsaw University. She works now on her doctorate and is engaged in psycho-visual tests, gamut-mapping and image quality measures.