*Journal of Imaging Science and Technology®* 54(3): 030201–030201-7, 2010.


# Image-Individualized Gamut Mapping Algorithms

**Zofia Barańczuk and Peter Zolliker**
*EMPA, Swiss Federal Laboratories for Materials Testing and Research, Dübendorf, Switzerland*
*E-mail: zofia.baranczuk@empa.ch*

**Joachim Giesen**
*Friedrich Schiller University in Jena, Jena, Germany*
*E-mail: joachim.giesen@uni-jena.de*

**Abstract.** *In this article the authors show that image quality measures can be successfully used to develop image-individualized gamut mapping algorithms. First the authors compare different image quality measures for the gamut mapping problem and then validate them using psychovisual data from four recent gamut mapping studies. The scoring function used to validate the quality measures is the hit rate, i.e., the percentage of correct choice predictions on data from psychovisual tests. Some of the image quality measures predict the observer's preferences as good as scaling methods such as Thurstone's method, which is used to evaluate the psychovisual tests. This is remarkable because the scaling methods are based on the experimental data, whereas the quality measures are independent of these data. The best performing image quality measure is used to automatically select the optimal gamut mapping algorithm for an individual image. © 2010 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.2010.54.3.030201]

## INTRODUCTION

Gamut mapping describes how a color image is rendered on a device with limited color reproduction capabilities. This classical problem is still an active area of research—Morovic gives a good recent overview.[1] An important step in improving a gamut mapping algorithm (GMA) is the accurate evaluation of its psychovisual performance. This is traditionally achieved using psychovisual tests, where observers have to decide which of alternative mappings is the best 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 study here is using an image quality measure (independent of observer feedback) to measure the perceived visual difference between a mapped image and the original. Image quality measures are successfully used in many imaging applications, such as modeling image distortions, especially in data compression.[3] An overview of the state-of-the-art image quality research can be found, for example, in Keelan[4] or Dijk.[5] The advantage of using image quality measures when evaluating gamut mapping algorithms is that they do not need additional psychovisual test data. Psychovisual tests generally give reliable results for the tested settings, but the tests are time consuming. Furthermore, an extrapolation to modified settings and new images is difficult. On the other hand computing an image quality measure provides results immediately. The challenge is to find a measure that correlates well with observers' preferences. 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 relevant; other factors encountered in different applications like noise or compression artifacts are of minor importance to gamut mapping.

The main topics of this article are exploring adequate image quality measures for comparing gamut mapping algorithms and the construction of image-individualized algorithms from a set of given nonindividualized gamut mapping algorithms by using adequate image quality measures. Different quality measures are evaluated by measuring their ability to predict observer choices in psychovisual test data.[6] General correlations of psychovisual gamut mapping evaluation and image quality measures have been studied before by Hardeberg[7] and Bonnier,[8] but here our focus is on predicting observers' choices in individual comparisons between mapped images. Image statistics were already used to improve the GMA of Bala,[9,10] where the window size for filtering was chosen based on the absolute value of the high pass filter output. Morovic[11] discussed which types of differences in images are most important for observers. Based on that, he built an image quality measure trying to predict observer's choices. Sun[12] analyzed a set of features extracted from original images to select an appropriate GMA from a range of GMAs. In this article we also try to find an optimal algorithm for individual images by using image quality measures to find the most similar mapped image compared to the given original image.

The remainder of this article is organized as follows: in the next section we describe the image quality measures considered in this article. Then Thurstone's method is briefly described as a method for evaluating psychovisual test data. In the subsequent section we describe how to evaluate the different image quality measures for gamut mapping. The data sets which we used for the evaluation are described in a separate section. Next, we discuss the experimental validation results on data sets. In the last section we use a particu-

Received Sep. 21, 2009; accepted for publication Mar. 18, 2010; published online May 3, 2010.
1062-3701/2010/54(3)/030201/7/$20.00.

lar image quality measure that turned out to perform well when used for constructing an image-individualized gamut mapping algorithm and compare its performance to the performance of nonindividualized reference gamut mapping algorithms.

## 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*

The structural similarity index (SSIM) has been introduced by Wang et al.[13] and is defined on quadratic image patches of size $k \times k$ at the same location within image $X$ and $Y$. 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 using the $L$ coordinate in CIELAB color space at the pixels:

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

$$\sigma^2_{P_X} = \frac{1}{k^2 - 1} \sum_{x \in P_X} (x - \bar{P}_X)^2,$$

$$\sigma^2_{P_Y} = \frac{1}{k^2 - 1} \sum_{y \in P_Y} (y - \bar{P}_Y)^2, \text{ and}$$

$$\sigma_{P_X P_Y} = \frac{1}{k^2 - 1} \sum_{i=1}^{k^2} (x_i - \bar{P}_X)(y_i - \bar{P}_Y). \tag{1}$$

The structural similarity index is then defined as

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

with two constants $c_1$ and $c_2$. As proposed by Wang et al.[3] 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 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 $Q_{\mathrm{SSIM}}$ value, the more similar are the compared images.

[†]In the literature, models describing distance or similarity between images are often called image quality metrics. Most of the measures described here are metrics according to the mathematical definition of this word, but not all of them (e.g. SSIM, ΔLC). Hence, as we want to have one word to describe all the models discussed here, we decided to use more general word "measure."

### *Laplacian Mean Square Error*

Like the SSIM the Laplacian mean square error (MSE) (LMSE), see Eskicioglu,[14] is a local measure for the difference between two images. We compute the following quantities on $L$ coordinate in CIELAB color space at each pixel, with indices $2 \leq i \leq n-1$ and $2 \leq j \leq m-1$) of $X$ and $Y$, respectively:

$$L(x_{ij}) = x_{(i+1)j} + x_{(i-1)j} + x_{i(j+1)} + x_{i(j-1)} - 4x_{ij} \tag{3}$$

and

$$L(y_{ij}) = y_{(i+1)j} + y_{(i-1)j} + y_{i(j+1)} + y_{i(j-1)} - 4y_{ij}. \tag{4}$$

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

$$Q_{\mathrm{LMSE}}(X, Y) = \frac{1}{(n-2)(m-2)} \sum_{i=2}^{n-1} \sum_{j=2}^{m-1} [L(x_{ij}) - L(y_{ij})]^2. \tag{5}$$

### *Mean Square Error*

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 $Q_{\mathrm{MSE}}$ is defined as

$$Q_{\mathrm{MSE}}(X, Y) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} (x_{ij} - y_{ij})^2, \tag{6}$$

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

### *Measure* $Q_{\Delta E}$

$\Delta E$ is defined as the Euclidean distance in CIELAB color space between corresponding pixels in two images $X$ and $Y$. That is, locally at pixel $x \in X$ and the corresponding pixel $y \in Y$ the $\Delta E$ distance is defined as

$$\Delta E(x, y) = \sqrt{[(L_x - L_y)^2 + (a_x - a_y)^2 + (b_x - b_y)^2]}. \tag{7}$$

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

$$Q_{\Delta E}(X, Y) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \Delta E(x_{ij}, y_{ij}). \tag{8}$$

$\Delta E$ is a popular image quality measure 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[15] or $\Delta E_{94}$.[16]

### *Measure* $Q_{\Delta \mathrm{LC}}$

The image quality measure $Q_{\Delta \mathrm{LC}}$ is based on a local contrast measure. We chose the Michelson contrast[17] 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 $k = 5$ in our experiments):

$$LC(P_X) = \frac{x_{max} - x_{min}}{x_{max} + x_{min}}, \quad (9)$$

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

$$\Delta LC(P_X, P_Y) = |LC(P_X) - LC(P_Y)|. \quad (10)$$

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

## THURSTONE'S METHOD

Traditionally image quality in gamut mapping is evaluated using Thurstone's Law of Comparative Judgment, which can be used to analyze paired comparison data (see Engeldrum[2]). Applying Thurstone's method allows us to derive a scale value for each tested gamut mapping algorithm. We use this values as a model to evaluate the quality of test images mapped with a gamut mapping algorithm. This model serves as a reference for the image quality measures discussed before, which do not need observer feedback in contrast to Thurstone's method.

To improve the consistency of results obtained by Thurstone's method, one can evaluate each image individually. Individualization linearly combines Thurstone's scale values for the entire data set with scale values obtained separately for each image. A description of this method can be found in Zolliker.[18] The coefficients in linear combination of those two scale values are then optimized on hold out data using cross validation. It turns out that scale values computed individually for images typically can be improved by shrinking them toward the scale values computed on the whole population of images—simply because in most cases there are not enough paired comparisons per image available. That is, shrinking provides a fall back when only a few paired comparisons are available for an image.

## EVALUATING THE QUALITY MEASURES

We want to assess the suitability of the different image quality measures for evaluating the quality of gamut mapping and compare it to Thurstone's method. Our validation procedure estimates how well the quality measures align with observers' ratings which we obtained in psychovisual tests. As mentioned before, the psychovisual test data is 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 by the percentage of correctly predicted observer choices, also known as hit rate. When computing hit rates for Thurstone's method we need to be careful that we do not validate the method on the same data that we used to derive the model—remember that Thurstone's method is, in contrast to the other image quality measures, based on observer data. To circumvent this problem we use cross validation, i.e., we use part of the data to learn a model and the remaining part to validate the model. In the following we provide more details on how to compute hit rates and how we employed cross validation.

### *Hit Rate*

For each paired comparison in a psychovisual test we know the choice of the observer. In some tests we allowed ties, i.e., neither of the two options is 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

$$HR = \frac{|S|}{|C|}, \quad (11)$$

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

### *Cross Validation*

We use cross validation to validate Thurstone's method. For this 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. Then each of the subsets is used once as testing set and the rest of the data as training set. The whole process is repeated ten times. The mean hit rate over all one hundred validation sets is used as the validation quality measure.

For the individualized variant of Thurstone's method, we carried out a double cross validation, i.e., 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 the optimizing set. Finally, we compute the hit rate on the validation set. We repeat this process 250 times and use the mean of the hit rates as validation quality measure.

## DATA SETS

The different image quality measures were validated on image data of four recent gamut mapping studies. All tests used paired comparisons, where two mapped images were compared to an original image. All those tests were carried out in a laboratory environment following the CIE guidelines.[19]

The first three studies used the ISO-Newspaper gamut as the target gamut, the fourth study used an OCE printer (OCE TCS 500) gamut. In the following we summarize the main ideas of the four studies.

### *Study 1: Basic Study*

This study is a traditional benchmark study comparing some recent image dependent gamut mapping algorithms to known reference algorithms. In addition to the reference algorithms HPminDE, SGCK,[19] the following algorithms us-

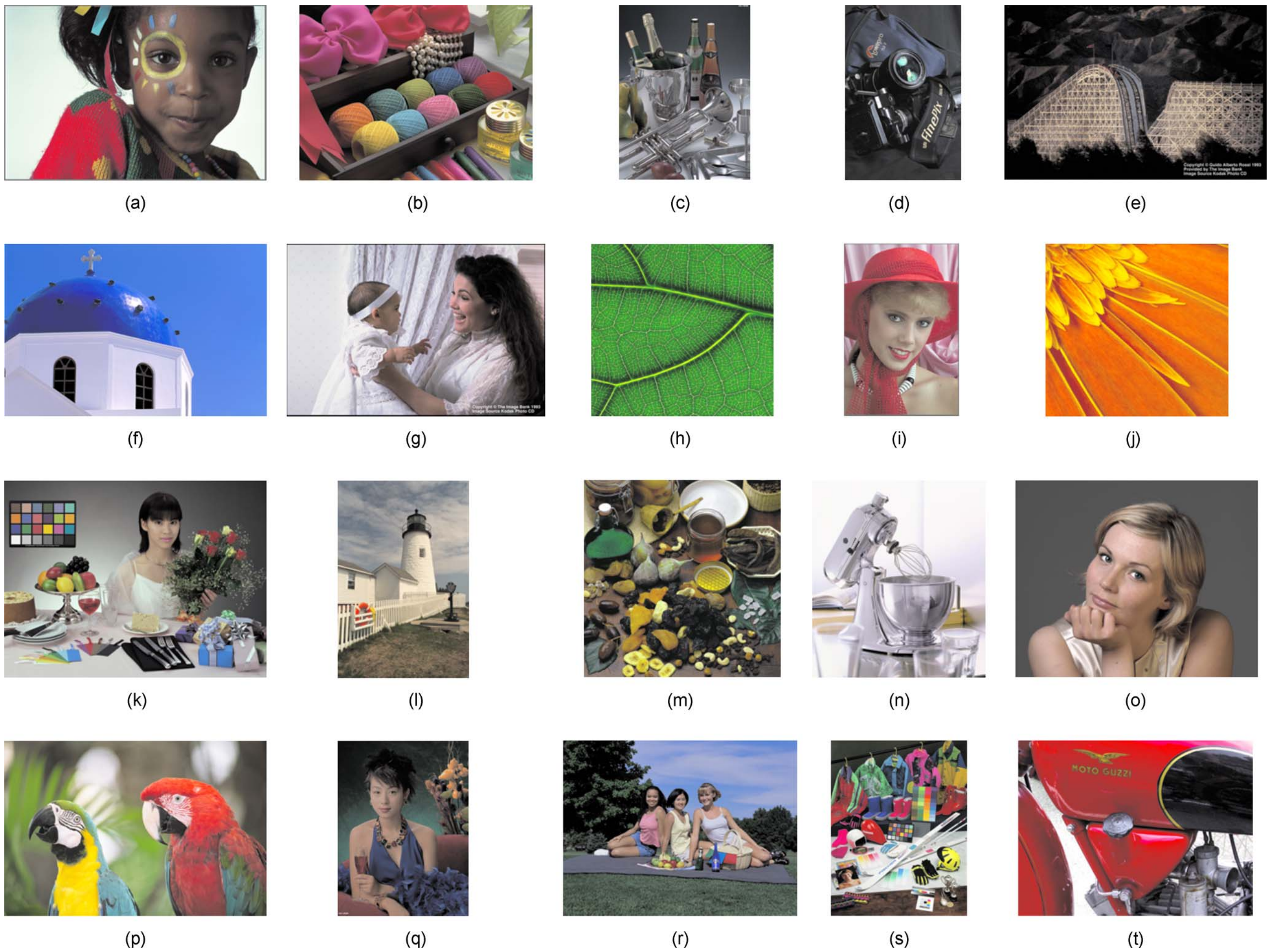


Figure 1. Images used in the IS study.

ing image gamut or spatial gamut mapping have been considered: NOptStar which is using the image gamut as described in Schuberth,[20] the Kolas algorithm,[21] the Caluori algorithm[22] and the Zolliker algorithm[23] applied to the SGCK and NOptStar algorithms. For this study 97 images were used, each mapped with seven algorithms. Each possible comparison, i.e., each pair of algorithms for each image, was done at least once. We will refer to this study as *basic study* or simply as *BS*.

### *Study 2: Image Gamut*

The topic of this study was the use of image gamut (IG) descriptions for gamut mapping.[20] 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 altogether in eight algorithms. Seventy-five images were used. Each possible comparison was done approximately twice. We will refer to this study as *image gamut study* or simply as *IG*.

### *Study 3: Local Contrast*

In this study[23] the influence of detail enhancement applied to a set of gamut mapping algorithms has been investigated. The study comprised HPminDE, SGCK, SGDA,[24] and a linear compression algorithm. All algorithms were compared with and without detail enhancement. Seventy-seven images were used, and 5376 comparisons have been performed. Each possible comparison was done approximately 2.5 times. We will refer to this study as *local contrast study* or simply as *LC*.

### *Study 4: Individual Study*

In this study[25,26] algorithms proposed by Gatta,[27] Kolas[21] and an algorithm using detail reconstruction proposed by Zolliker[23] applied to HPMinDE were compared with the reference algorithms HPminDE and SGCK. Twenty images, presented in Figure 1, have been used. Each possible comparison was done 40 times. We will refer to this study as *individual study* or simply as IS.

## EXPERIMENTAL RESULTS

The results we are discussing here are hit rates obtained by using the different image quality measures and Thurstone's method on the four psychovisual test data sets. The hit rates are summarized in Figure 2.

In the following we will first review these hit rates in more detail, and then discuss the theoretical upper limit of the hit rate for the IS study.

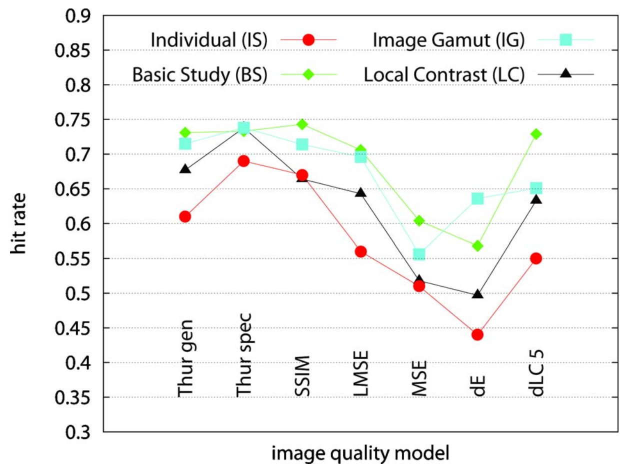


Figure 2. Hit rates obtained by different methods for four studies. Here *Thur gen* refers to general Thurstone's method and *Thur spec* refers to individualized Thurstone's method.

### *Evaluating Thurstone's Method*

In Figure 3 we present the hit rates for Thurstone's method on the different data sets and different degrees of individualization. As expected, hit rates on training sets are higher than those on test sets. Individualization improves training set hit rates, however it does not always improve hit rates on the test sets. The higher hit rate on the training set is due to the specialization of the model to the data set. For a conclusion to general data the hit rates on the test set have to be used.

On the *BS* data set individualization does not increase the hit rate on the test sets. These test sets included only about one repetition of each comparison, so individual results are probably not stable enough to contribute to the model accuracy.

For the IG study the optimal hit rate is obtained for a linear combination of the global scale values and individualized ones. In this test each comparison was repeated twice, which is enough to individualize the scale values but not enough to get a significantly better hit rate for these scale values than for the global scale values. The best hit rate needs a combination of global and individualized scale values.

In the LC study there are about 2.5 repetitions for each comparison. As for the *IG* study, the optimal hit rate is obtained for a combination of individual and global scale values. But in the LC study using only the individual scale values provides a hit rate almost as high as the optimal combination of global and individualized scale values.

The largest number of repetitions for the individual comparisons between algorithms, namely, 40, is in IS study. Here, the highest hit rate is obtained using just the individual scales values.

### *Evaluating the Image Quality Measures*

On all the data sets the structural similarity (SSIM) measure proved to be the best performing image quality measure, i.e., it achieved the highest hit rates. On the BS data set the results obtained with SSIM are even better than those coming from individualized Thurstone's method. In the other studies the individualized Thurstone's method yields better results. As discussed before, a likely reason for this behavior is the size of the BS study as the performance of Thurstone's method improves with an increasing number of comparisons. It is worth noting that the hit rates obtained for SSIM are comparable to the hit rates for the general Thurstone's method or, in case of the IS test, even much higher.

The two pointwise image quality measures that we considered, namely, $Q_{\Delta E}$ and the mean square error $Q_{\mathrm{MSE}}$, scored lower than their competitors, often showing hit rates close to random choice, i.e., 50%. The likely reason is that all gamut mapping algorithms tested in these studies already optimize color preservation in some way, and thus observers' choices are more affected by detail preservation. In particular, clipping algorithms, for example, HPminDE, are optimizing the mapped image against pointwise distance measures but ignore detail preservation.

The quality measures LMSE and LC, which embody details preservation differences, perform better than pointwise measures but still not as good as SSIM.

### *Theoretical Limit Hit Rates*

The theoretical limit hit rate of 1.0 is almost never achieved because observers usually differ in their choices and even the decisions of a single observer are typically inconsistent, i.e., the same person, under the same conditions makes a different choice on the same images in a repeated paired comparison, if the choice is difficult.

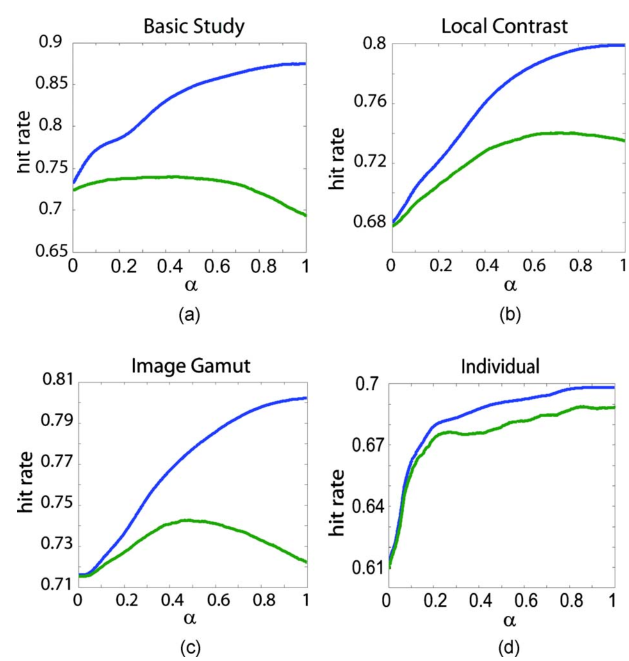


Figure 3. Hit rates using Thurstone's method for: (a) basic study, (b) local contrast study, (c) image gamut study, and (d) individual study. The higher line shows the hit rate on the training set, the lower line shows the hit rate on the test set. Scale values ($sv$) are computed as a convex combination of scale values for the whole population of images ($sv_{\mathrm{group}}$) and scale values for individual images ($sv_{\mathrm{ind}}$), i.e., $sv=\alpha sv_{\mathrm{ind}}+(1-\alpha)sv_{\mathrm{group}}$ with $\alpha>0$.

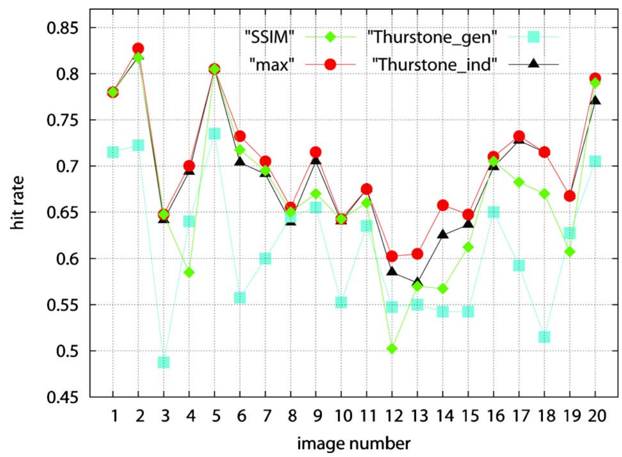


Figure 4. Hit rates obtained by different methods for the different images in the IS test set. Again, Thur gen refers to Thurstone's method without individualization, and Thur spec refers to Thurstone's method without individualization.

If we have choice data with many repetitions for each choice, then we can estimate a better (choice data dependent) limit for the hit rate than the ideal 1.0, namely, the *maximally achievable hit rate* as follows: let $F_{ij}$ be the frequency that algorithm $i$ has been preferred over algorithm $j$ in comparison, i.e., the number of times an image mapped using algorithm $i$ has been preferred over the same image mapped by algorithm $j$ divided by the total number $i$ and $j$ have been compared. If we have the same number of repetitions for each comparison (which was the case in the IS test), we can define the maximal achievable hit rate as follows:

$$HR_{\max} = \frac{\sum_{i<j} \max(F_{ij}, F_{ji})}{\text{number of algorithm pairs}}. \qquad (12)$$

In this article we considered maximal achievable hit rates per image. Since we have many repetitions for each comparison in the IS study data set, we use this data set to check how close the best performing quality measures SSIM and Thurstone's method come to the maximally achievable hit rate. The hit rates computed for the different images in the IS test set are shown in Figure 4.

The hit rates obtained using Thurstone's method with individualization is always very close to the maximally achievable hit rate for all images. For many images the two hit rates are even equal. The hit rates achieved by SSIM are lower, but generally close to the ones for Thurstone's method with individualization and much higher than for Thurstone's method without individualization. Only on three images out of 20 SSIM performs worse than Thurstone's method without individualization.

## USING SSIM TO CONSTRUCT AN IMAGE-INDIVIDUALIZED GAMUT MAPPING ALGORITHM

The results from the previous sections suggest that we can design a meta gamut mapping algorithm that chooses a "best" gamut mapping for a given image from a class of mappings. Here best is meant with respect to an image quality measure that proved to be well suited to predict the perceived quality of a mapping. Again, the previous sections suggest that SSIM is suitable as such a measure. This approach is also supported by previous studies[1,12,25] showing that different gamut mapping algorithms perform differently on different images, i.e., one can improve the quality of the mapped images by choosing the best algorithm for each image, instead of using the same algorithm for all images.

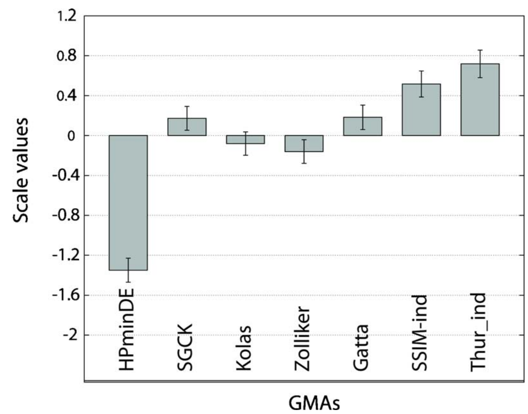


Figure 5. Scale values of five algorithms considered in the IS study, plus the meta-algorithm based on the SSIM measure and Thurstone's individualized scale values. Error bars show one standard deviation and are computed as in Zolliker.[18]

Formally, the meta-algorithm can be described as follows: for a given quality measure $Q$ (in our case SSIM) and a given image $I$, let $I_1, \ldots, I_n$ be the mappings of this image using $n$ different mapping algorithms. Choose the mapping $I_k$ such that $Q(I_k) \geq Q(I_i)$ for all $i = 1, \ldots, n$.

### *Validation of the Meta-algorithm*

Thurstone's method can easily be adapted to compare the quality of the meta-algorithm and the individual algorithm on which the meta-algorithm builds. We used the data from the IS study to validate two meta-algorithms, one using SSIM, another using scale values from the individualized Thurstone's method on the training set as quality measure. Remember that this study comprised twenty images, each of them mapped by five algorithms. In Figure 5 we summarize the results of the comparison.

The image-individualized algorithm based on Thurstone's method performs significantly better than any single algorithm. The meta-algorithm using SSIM does not quite reach the scale value of Thurstone's meta algorithm, but has higher scale values than the best single algorithm. The differences between SSIM-optimized algorithm and the above mentioned algorithms are not very significant, hence more testing is needed. However, the algorithm based on the individualized Thurstone's method is not practical, as it requires conducting a psychovisual test for every image we want to map. Still, also the SSIM measure has its limitations, e.g., as can be seen in Fig. 4 the meta-algorithm predicts choices for image number 4, 12, or 19 worse than Thurstone's general scale values. Note that we applied SSIM only for the $L$-coordinate of CIELAB space, thus image quality effects based on the color coordinates have been neglected. The meta-algorithm based on a generalized SSIM-like quality measure using all three color space dimensions may perform even better.

## CONCLUSIONS

We demonstrated that image quality measures can be a useful and efficient method to gauge the quality of mapped images in gamut mapping. In our study, 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. Moreover, we have shown that image quality measures such as SSIM can successfully be used to select the optimal algorithm for a given image.

## ACKNOWLEDGMENTS

We thank the Norwegian Color Research Laboratory for making their psychovisual test data available. This work was supported by the Hasler Stiftung Project No. 2200.

## REFERENCES

[1] J. Morovic, *Color Gamut Mapping* (WileyBlackwell, Chichester, UK, 2008), ISBN: 0470030321.

[2] P. G. Engeldrum, *Psychometric Scaling: A Toolkit for Imaging Systems Development* (Imcotek Press, Winchester, MA, 2000).

[3] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity", IEEE Trans. Image Process. **13**, 600 (2004).

[4] B. Keelan, *Handbook of Image Quality* (Marcel Dekker, New York, 2002).

[5] J. Dijk, "In search of an objective measure for the perceptual quality of printed images", Ph.D. thesis, Delft University of Technology, 2004.

[6] Z. Baranczuk, P. Zolliker, and J. Giesen, "Image quality measures for evaluating gamut mapping", Proceedings IS&T/SID 17th Color Imaging Conference. (IS&T, Springfield, VA, 2009), pp. 21–26.

[7] J. Y. Hardeberg, E. Bando, and M. Pedersen, "Evaluating color image difference metrics for gamut mapping images", Color. Technol. **124**, 243 (2008).

[8] N. Bonnier, F. Schmitt, H. Brettel, and S. Berche, "Evaluation of spatial gamut mapping algorithms", *Proc. IS&T/SID 14th Color Imaging Conference* (IS&T, Springfield, VA, 2006), pp. 56–61.

[9] R. Bala, R. deQueiroz, R. Eschbach, and W. Wu, "Gamut mapping to preserve spatial luminance variations", J. Imaging Sci. Technol. **45**, 436 (2001).

[10] R. Bala, R. deQueiroz, R. Eschbach, and W. Wu, "Gamut mapping to preserve spatial luminance variations", Proc. IS&T/SID 8th Color Imaging Conference (IS&T, Springfield, VA, 2000) pp. 122–128.

[11] J. Morovic and P. L. Sun, "Predicting image differences in color reproduction from their colorimetric correlates", J. Imaging Sci. Technol. **47**, 509 (2003).

[12] P. L. Sun and Z. W. Zheng, "Selecting appropriate gamut mapping algorithms based on a combination of image statistics", Proc. SPIE **5667**, 211–219 (2004).

[13] Z. Wang and A. Bovik, "A universal image quality index", IEEE Signal Process. Lett. **9**, 81 (2002).

[14] A. Eskicioglu and P. Fisher, "Image quality measures and their performance", IEEE Trans. Commun. **43**, 2959 (1995).

[15] N. Moroney, M. Fairchild, R. Hunt, C. Li, M. Luo, and T. Newman, "The CIECAM02 color appearance model", *Proc. IS&T/SID 10th Color Imaging Conference* (IS&T, Springfield, VA, 2002) pp. 23–27.

[16] R. McDonald and K. Smith, "CIE94: A new color-difference formula", J. Soc. Dyers Colour. **111**, 376 (1995).

[17] A. A. Michelson, *Studies in Optics* (Univ. Chicago Press, Chicago, IL, 1927).

[18] P. Zolliker, Z. Baranczuk, I. Sprow, and J. Giesen, "Conjoint analysis used for the evaluation of parameterized gamut mapping algorithms", IEEE Trans. Image Process. **19**, 758 (2010).

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

[20] J. Giesen, E. Schuberth, K. Simon, and P. Zolliker, "Image-dependent gamut mapping as optimization problem", IEEE Trans. Image Process. **16**, 2401 (2007).

[21] O. Kolas and I. Farup, "Efficient hue-preserving and edge-preserving spatial color gamut mapping", Proc. IS&T/SID 15th Color Imaging Conference Final Program and Proceedings (IS&T, Springfield, VA, 2007) pp. 207–209.

[22] U. Caluori and K. Simon, "An RGB color management concept based on an improved gamut mapping algorithm", Proc. SPIE **7241**, 724118A (2009).

[23] P. Zolliker and K. Simon, "Retaining local image information in gamut mapping algorithms", IEEE Trans. Image Process. **16**, 664 (2007).

[24] P. Zolliker and K. Simon, "Continuity of gamut mapping algorithms", J. Electron. Imaging **15**, 013004 (2006).

[25] F. Dugay, "Perceptual evaluation of color gamut mapping algorithms", M.Sc. thesis, The Norwegian Color Research Laboratory, 2007.

[26] F. Dugay, I. Farup, and J. Hardeberg, "Perceptual evaluation of color gamut mapping algorithms", Color Res. Appl. **33**, 470 (2008).

[27] I. Farup, C. Gatta, and A. Rizzi, "A multiscale framework for spatial gamut mapping", IEEE Trans. Image Process. **16**, 2423 (2007).