Objective evaluation of naturalness, contrast, and colorfulness of tone-mapped images

Lukáš Krasula, Karel Fliegel, Patrick Le Callet, Miloš Klíma

Objective evaluation of naturalness, contrast, and colorfulness of tone-mapped images

Lukáš Krasula[a,b], Karel Fliegel[a], Patrick Le Callet[b], and Miloš Klíma[a]

[a]Czech Technical University in Prague, Technická 2, 166 27 Prague 6, Czech Republic;
[b]LUNAM University - IRCCyN CNRS UMR 6597, 44306, Nantes, France

ABSTRACT

The main obstacle preventing High Dynamic Range (HDR) imaging from becoming standard in image and video processing industry is the challenge of displaying the content. The prices of HDR screens are still too high for ordinary customers. During last decade, a lot of effort has been dedicated to finding ways to compress the dynamic range for legacy displays with simultaneous preservation of details in highlights and shadows which cannot be achieved by standard systems. These dynamic range compression techniques are called tone-mapping operators (TMO) and introduce novel distortions such as spatially non-linear distortion of contrast or naturalness corruption. This paper provides an analysis of objective no-reference naturalness, contrast and colorfulness measures in the context of tone-mapped images evaluation. Reliable measures of these features could be further merged together into single overall quality metric. The main goal of the paper is to provide an initial study of the problem and identify the potential candidates for such a combination.

Keywords: High Dynamic Range (HDR), Tone-Mapping Operators (TMO), Image Quality Measures, Objective Evaluation

1. INTRODUCTION

Most of the digital image and video applications work with 8-bits per color channel. The dynamic range of real world scenes, which is usually much higher, is therefore compressed causing the presence of underexposed and overexposed areas in images and videos. High Dynamic Range (HDR) imaging brings ways to capture the whole range of luminance values in the scene. Currently, there are several possibilities for HDR image acquisition,[1] including usage of cameras with special HDR or multiple chips. However, the most common procedure exploits the exposure bracketing. The scene is captured multiple times with different exposure settings. The particular images are then fused together creating a radiance map of the scene. The main advantage of this approach is its universality, since it enables the use of virtually any camera. The problems occur when the scene being captured is dynamic. The shifts among the images has to be handled by post-processing and even so, mostly some motion artifacts appear. A problem with hand-shake can be resolved by a tripod but this can also limit a photographer's comfort. Another possibility could be an exploitation of a photographic film's higher dynamic range. The negative could be scanned with different exposure times and processed in the similar way as mentioned.

Nevertheless, the acquisition of an HDR content is not the most crucial problem in the area of HDR imaging. The real challenge is displaying it. Despite the existence of technologies for HDR screens, their availability on consumer market is still highly limited, not mentioning their high price and excessive weight and thickness. For enabling the use of common low dynamic range (LDR) screens, Tone-Mapping Operators (TMO) have been invented. These algorithms are specifically designed for compressing the dynamic range of an image or a video while simultaneously preserving the details in highlights and shadows provided by HDR acquisition. TMO which apply the same compression function to each pixel are referred to as global TMO. Local TMO, on the other hand, adapt the function according to the pixel neighborhood. This allows them to reproduce more details but

Further author information: (Send correspondence to L. Krasula)
L. Krasula: E-mail: krasuluk@fel.cvut.cz, lukas.krasula@univ-nantes.fr
K. Fliegel: E-mail: fliegek@fel.cvut.cz
P. Le Callet: E-mail: patrick.lecallet@univ-nantes.fr
M. Klíma: E-mail: klima@fel.cvut.cz

make them more likely to corrupt the overall naturalness of the scene and to introduce undesirable artifacts. The computational costs are also significantly higher. A number of studies have been conducted in order to compare the particular TMO and to decide which of the two approaches provides better results. Some of them are summarized in section 1.1.

For most of the practical applications of HDR and tone-mapping, objective algorithms for quality assessment are needed. The effort dedicated to the development of such methods is described in section 1.2.

This paper focuses on assessing some of the most dominant features of the tone-mapped images influencing the final quality as identified by Čadík et al.[2] Namely estimators of naturalness (section 2.4), contrast (section 2.2), and colorfulness (section 2.3) are tested against the subjective ranks provided in the mentioned paper. The goal is to identify the best candidates for merging into a single no-reference overall tone-mapped images quality measure. In section 3, the conclusions are provided and future work is discussed.

1.1 Subjective Studies for TMO Evaluation

Since the outcomes of TMO introduce different artifacts and distortions, subjective studies were carried out to provide an insight into observers' perception and preferences. One of the first were conducted by Drago et al.[3] Here, the paired comparison procedure has been employed. Two most important factors influencing the overall quality were identified as naturalness and details.

The pair comparison was used also by Kuang et al.[4] The paper is a summary of several experiments conducted by the group. Subjects were supposed to decide between the pairs based on overall quality and other features. This enabled to conclude, which attributes contribute the most to the resulting quality. Overall contrast, colorfulness, sharpness and details in shadows seemed to correlate well with the preferences.

Yoshida et al.[5] compared the global and local TMO in terms of perceptual attributes such as naturalness, contrast, details, etc. As expected, local TMO outperformed the global ones in reproduction of details but seemed to be worse in handling contrast and brightness.

The study by Ledda et al.[6] utilized an HDR display as a reference. Observers were then supposed to select the closer image from simultaneously displayed LDR image pair.

Subjective perception of HDR images displayed on screens with different dynamic range and as a real-world scenes was studied by Yoshida et al.[7]

The works by Ashikhmin and Goyal[8] and Čadík et al.[2] both proved that there is no significant difference between observers' preferences once the real-world reference is provided. However, the results of the fidelity tests can differ. Moreover, in the paper by Čadík[2] subjects' preferences about overall quality and various image attributes are provided for numerous TMO.

Akyüz et al.[9] showed in their experiments that the same content displayed on the HDR display is mostly preferred over LDR screen, even though the simplest methods are used for LDR to HDR conversion. They also discovered that in most of the cases, LDR image with optimal exposure is perceived as the same or better than the tone-mapped HDR image.

A different approach was chosen by Barkowsky and Le Callet.[10] The purpose of the study was to test the abilities of particular TMO. Three parameters – contrast, saturation and details – were varied within two TMO. Observers selected the best looking combination of the parameters using hardware sliders. Thus the outcomes of the TMO were optimized.

More quantitative studies providing different TMO comparisons were done e.g. by Kuang et al.,[11] Song et al.,[12] or Yeganeh and Wang.[13]

1.2 Objective Metrics for TMO Evaluation

The problem of objective quality assessment of HDR images was firstly addressed in 2005 by Mantiuk et al.[14] They used the concept of Visual Detection Predictor (VDP) developed by Daly.[15] However, it was designed only for the comparison of two images with the same dynamic range and therefore can not be used for evaluation of tone-mapped images.

In 2008, Aydin et al.[16] introduced the Dynamic Range Independent Metric (DRIM) which compares the images regardless the difference in their dynamic ranges. It addresses three features - loss of visible contrast (when the contrast is visible in reference but not in the processed version), amplification of invisible contrast (opposite case to the previous one), and reversal of visible contrast (the contrasts are visible in both of the images but have different polarity). The output of this metric are three distortion maps. Unfortunately, the algorithm does not provide the estimation of quality as a single value. The pooling of the distortion maps requires further analysis.

Mantiuk et al. have revised the HDR-VDP concept in 2011, resulting in the novel HDR-VDP-2 metric.[17] Regardless the name, the metric is fundamentally different from its ancestor. Unlike the previously mentioned algorithms, HDR-VDP-2 also tries to provide a single value of overall quality. This is however calibrated for standard LDR images. More meaningfull measure, also provided by the metric, seems to be the *visibility*, which quantifies the visible differences between the HDR images.

The only metric designed specifically for the evaluation of tone-mapped images is Tone-Mapped Image Quality Index (TMQI) proposed by Yeganeh and Wang in 2013.[13] It combines the structural similarity measure with the natural scene statistics (NSS). The structural similarity is calculated similarly to the multi-scale SSIM[18] but adjusted to the comparison of images with different dynamic range. NSS metric is applied on the tone-mapped image only.

2. PARTICULAR FEATURES ESTIMATION

As summarized above, a lot of effort has been dedicated to investigate how overall quality is influenced by various attributes. Čadík et al.[2] have also performed the fusion of features showing their particular contributions. If we are able to reliably estimate these attributes, it should be possible to find a merging function that would provide a good measure of overall quality. In this work three features, identified as key players in most of the studies, are considered. These are naturalness, contrast, and colorfulness of an image.

2.1 Methodology for Testing the Performance of the Measures

For the performance evaluation of particular measures, correlation with ranks provided by human subjects is calculated. Two well known non-parametric correlation coefficients are used – Kendall's tau (KRCC) and Spearman's rho (SROCC). The first is computed as

$$KRCC = \frac{N_c - N_d}{N}, \tag{1}$$

where N_c represents the number of concordant pairs in the series (i.e. the pairs which have the same order in both series), N_d is the number of discordant pairs (the pair's order differs within the series), and N is the total number of image pairs. If n is the number of images in a set, $N = \frac{1}{2}n(n-1)$.

In Spearman's rank order correlation coefficient calculation, firstly the particular values are transformed into ranks for both series. Then

$$SROCC = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}, \tag{2}$$

where d_i is the difference of ranks for i-th image and n is the number of images in a set.

For contrast and colorfulness estimators the rank's provided by Čadík et al.[2] in their *ranking* experiment for these particular attributes were used as ground-truth. There are three source contents in the dataset tone-mapped by 14 TMO evaluated by 10 observers.

Since the ranks according to naturalness only are not provided, it could be tested solely in terms of correlation with the overall quality. However, as concluded by the above mentioned studies, this correlation is perceived as very high by human subjects. Moreover, the subjects in Čadík's experiment were instructed to provide ranks according to their mental representation of the real world scene. For more diversity, also the dataset by Yeganeh and Wang[13] was employed. Here, only overall quality ranks obtained from 20 subjects for 15 source contents and 8 TMO are available and thus the results for naturalness estimators could be slightly biased by the small number of cases where the image might be more appealing, even though it is less natural.

2.2 Contrast Estimation

Contrast is one of the most important features of any image. A lot of work has been dedicated to contrast perception and measurement. An overview of contrast estimators for grayscale and color images was elaborated by Panetta et al.[19] In this paper, measures are applied on the luminance component of images only and the size of each block is set to be 11 $\times$ 11 pixels.

The first measure is based on the Weber's law, which is defined for uniform background. It was formulated as a measure of enhancement by Agaian et al.[20] which is calculated as

$$C_{\text{weber}} = \frac{1}{B}\sum_{b=1}^{B} 20 \ln \frac{I_{\max,block_b}}{I_{\min,block_b}}, \tag{3}$$

where B is the number of blocks an image is divided to, $I_{\max,block_b}$ and $I_{\min,block_b}$ are the maximal and the minimal intensity value within the block b. The same notation is maintained also hereafter.

The second measure employs a concept of Michelson contrast, adjusted by Agaian et al.[21] The computation is

$$C_{\text{michelson}} = -\frac{1}{B}\sum_{b=1}^{B} 20 \ln \frac{I_{\max,block_b} - I_{\min,block_b}}{I_{\max,block_b} + I_{\min,block_b}}. \tag{4}$$

A metric less sensitive to noise is Second Derivative-based Measure of Enhancement, proposed by Panetta et al.[22] It directly connects each pixel to the contrast value which can be exploited in various ways, e.g. for image contrast enhancement.[23] It is implemented as

$$C_{\text{sdme}} = -\frac{1}{B}\sum_{b=1}^{B} 20 \ln \left| \frac{I_{\max,block_b} - 2I_{\text{center},block_b} + I_{\min,block_b}}{I_{\max,block_b} + 2I_{\text{center},block_b} + I_{\min,block_b}} \right|, \tag{5}$$

where $I_{\text{center},block_b}$ is the intensity of the center pixel of each block (i.e. the pixel under consideration).

Another widely exploited concept is RMS contrast. It is defines as

$$C_{\text{rms}} = \sqrt{\frac{1}{B}\sum_{b=1}^{B} \left(I_{\text{center},block_b} - I_{\text{mean},block_b}\right)^2}, \tag{6}$$

where $I_{\text{mean},block_b}$ is the mean intensity of the block b.

This was further elaborated by Panetta et al.[19] Properties of human visual system (HVS) were integrated and the original was modified to Root Mean Enhancement measure, calculated as

$$C_{\text{rme}} = \frac{1}{B}\sqrt{\sum_{b=1}^{B} \left| \frac{\ln |I_{\text{center},block_b} - I_{\text{mean},block_b}|}{\ln |I_{\text{center},block_b} + I_{\text{mean},block_b}|} \right|}. \tag{7}$$

A different approach has been proposed by Matković et al.[24] They firstly approximate the perceptual luminance $L(i,j)$ of pixels by

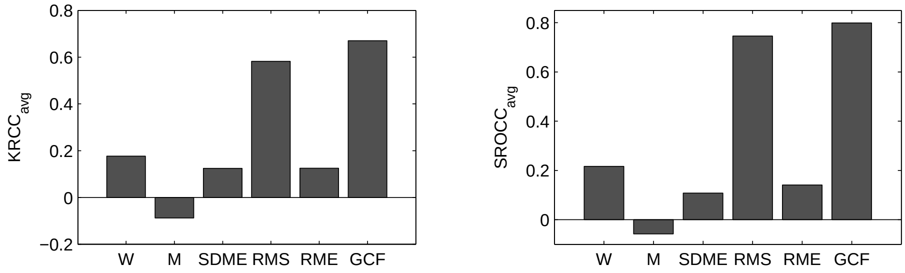
$$L(i,j) = 100\sqrt{\left(\frac{I(i,j)}{255}\right)^{2.2}}, \tag{8}$$

where $I(i,j)$ is the original intensity of the pixel in the position i, j. Then for every pixel the average difference between itself and its neighbors above, below, on the left, and on the right is calculated. These average differences are then averaged again, providing the single contrast value C_k. This procedure is repeated for nine resolutions. The final Global Contrast Factor (GCF) is obtained by weighting

$$GCF = \sum_{k=1}^{9} w_k C_k, \tag{9}$$

Scene	Weber		Michelson		SDME		RMS		RME		GCF	
	KRCC	SROCC	KRCC	SROCC	KRCC	SROCC	KRCC	SROCC	KRCC	SROCC	KRCC	SROCC
1	-0.1547	-0.3168	0.0884	0.2156	-0.0221	-0.1650	0.3315	0.5655	-0.3315	-0.4884	0.4199	0.5633
2	0.3222	0.5330	0.0556	0.0705	-0.0111	-0.0308	0.6778	0.7908	0.3000	0.4427	0.8334	0.9339
3	0.3626	0.4330	-0.4066	-0.4593	0.4066	0.5209	0.7363	0.8813	0.4066	0.4681	0.7582	0.8989

Table 1. Performance of the selected contrast measures on the dataset developed by Čadík et al.[2] with ranks according to subjectively perceived contrast.



(a) Average KRCC (b) Average SROCC

Figure 1. Average correlation coefficients of selected contrast measures.

where w_k is the weight of the particular resolution k. The authors propose to calculate the weights from

$$w_k = (-0.406385\frac{k}{9} + 0.334573)\frac{k}{9} + 0.0877526. \tag{10}$$

The performance of the above described criteria quantified in terms of Kendall's and Spearman's correlation coefficients with Čadík's[2] ranks according to contrast is depicted in Table 1. The best results are highlighted in gray. For better illustration the average correlation coefficients are depicted as a bar graphs in Figure 1.

It can be seen, that most of the measures do not behave systematically well, i.e. they do not even provide the same mark of the coefficient for different scenes. The exceptions are RMS contrast and GCF which on the other hand correlate very well with the subjective scores for scenes #2 and #3. The scene #1 seems to be more challenging. Nevertheless, RMS contrast and GCF could be possible candidates for incorporation into the overall quality metric.

2.3 Colorfulness Estimation

Incorporating estimation of colorfulness into quality assessment has been extensively studied e.g. by Hasler and Suesstrunk[25] or Panetta et al.[19] In this section three perspective measures are identified and tested in the given context.

The first potential candidate is CIQI colorfulness inspired by Hasler and Suesstrunk[25] and modified by Fu.[26] It is defined as

$$CIQI_c = \big(\sqrt{\sigma_\alpha^2 + \sigma_\beta^2} + 0.3\sqrt{\mu_\alpha^2 + \mu_\beta^2}\big)/85.59, \tag{11}$$

where

$$\mu_\alpha = \frac{1}{N}\sum_{p=1}^{N}\alpha_p = \frac{1}{N}\sum_{p=1}^{N}(R_p - G_p), \tag{12}$$

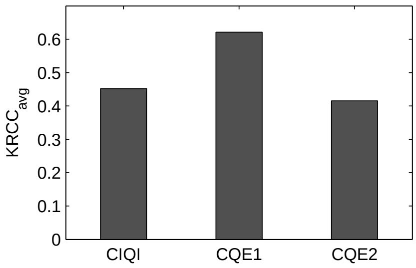
$$\mu_\beta = \frac{1}{N}\sum_{p=1}^{N}\beta_p = \frac{1}{N}\sum_{p=1}^{N}\big(0.5(R_p + G_p) - B_p\big), \tag{13}$$

$$\sigma_\alpha^2 = \frac{1}{N}\sum_{p=1}^{N}(\alpha_p^2 - \mu_\alpha^2). \tag{14}$$

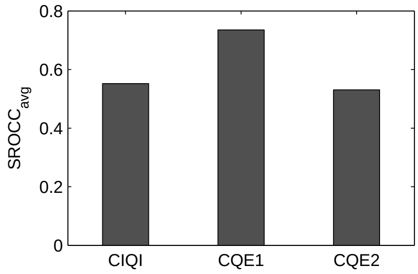
N is the number of pixels, p is the pixel index (considering the image is put into a single column vector), and R, G, B are red, green and blue component of an image. The calculation of σ_β^2 is similar to (14).

Scene	CIQI$_c$		CQE1$_c$		CQE2$_c$	
	KRCC	SROCC	KRCC	SROCC	KRCC	SROCC
1	0.8398	0.9483	0.8840	0.9615	0.7072	0.8801
2	-0.0663	-0.0506	0.3094	0.4026	-0.0221	-0.0330
3	0.5824	0.7582	0.6703	0.8418	0.5604	0.7451

Table 2. Performance of the selected colorfulness measures on the dataset developed by Čadík et al.[2] with ranks according to color representation.



(a) Average KRCC



(b) Average SROCC

Figure 2. Average correlation coefficients of selected colorfulness measures on.

The whole concept was further elaborated by Panetta et al.[19] who introduced properties of HVS into computations. Two other estimators are defined - CQE1 colorfulness and CQE2 colorfulness. The first is implemented as

$$CQE1_c = 0.2 \times \ln\left(\frac{\sigma_\alpha^2}{|\mu_\alpha|^{0.2}}\right) \times \ln\left(\frac{\sigma_\beta^2}{|\mu_\beta|^{0.2}}\right), \tag{15}$$

and the second as

$$CQE2_c = 0.2 \times \frac{\ln\sigma_\alpha^2 \ln\sigma_\beta^2}{\ln\sigma_c^2} \times \frac{\ln\mu_\alpha^2 \ln\mu_\beta^2}{\ln\mu_c^2}, \tag{16}$$

where

$$\mu_c = \frac{1}{2N}\left(\sum_{p=1}^{N}\alpha_p + \sum_{p=1}^{N}\beta_p\right), \tag{17}$$

and

$$\sigma_c^2 = \frac{1}{2N}\left(\sum_{p=1}^{N}(\alpha_p^2 - \mu_c^2) + \sum_{p=1}^{N}(\beta_p^2 - \mu_c^2)\right). \tag{18}$$

The performance of the above described colorfulness estimators was again evaluated using Kendall's and Spearman's correlation coefficient. This time the ranks according to color representation provided by Čadík et al.[2] were considered. The results are depicted in the Table 2. The average correlation coefficients are shown in Figure 2.

The only estimator having consistently positive correlation coefficients is $CQE1_c$. Thus it seems to be most suitable for further usage from the tested criteria. The most challenging scene for colorfulness measures is #2 (unlike for the contrast estimators where it was scene #1).

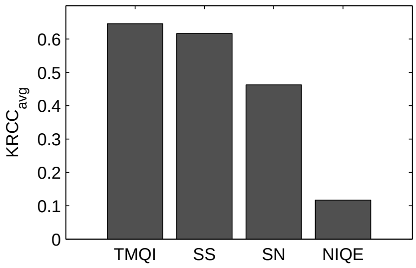
2.4 Naturalness Estimation

The statistics of natural scenes (NSS) have been a research matter for a long time and a number of models have been developed.[27,28] In the field of image quality assessment it has been incorporated in full-reference[29] as well as no-reference[30] metrics. Here, two naturalness estimators are selected and tested. The particular selection was guided by their invariance toward the type of distortion and no need for reference.

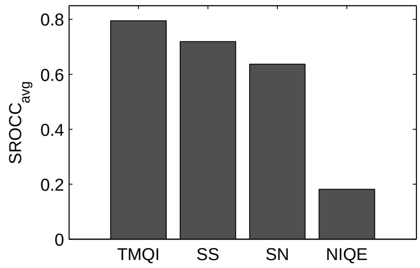
The first measure tested is incorporated in full-reference TMQI metric.[13] It uses the concept of modeling the brightness and contrast distributions in natural images. Namely, the histograms of means and mean standard deviations for 3000 images were computed and fitted by the Gaussian and Beta function, respectively. Based

Scene	TMQI		SS		SN		NIQE	
	KRCC	SROCC	KRCC	SROCC	KRCC	SROCC	KRCC	SROCC
1	0.5304	0.7063	0.5304	0.6931	0.4199	0.5919	-0.1326	-0.1188
2	0.6703	0.8418	0.4286	0.4901	0.3407	0.5297	0.3846	0.5209
3	0.7363	0.8374	0.8901	0.9736	0.6264	0.7890	0.0989	0.1429

Table 3. Performance of quality and naturalness measures on Čadík's dataset.[2]



(a) Average KRCC



(b) Average SROCC

Figure 3. Average correlation coefficients of quality and naturalness measures on Čadík's dataset.[2]

on the assumption of brightness and contrast independence, their normalized joint probability density function (PDF) is their multiple and the Statistical Naturalness (SN) measure is defined as

$$SN = \frac{1}{K} P_m P_d, \tag{19}$$

where K is the normalization factor equal to $\max\left[P_m, P_d\right]$.

The second estimator was developed by Mittal et al.[31] and is called Natural Image Quality Evaluator (NIQE). It is based on the model developed by Ruderman.[28] After the preprocessing of an image and salient areas identification, 18 'quality aware' NSS features are calculated on two scales. These are then modeled using Multivariate Gaussian model (MVG) and compared to the model obtained from 125 natural images. Since the MVG are represented by their mean vectors ν and covariance matrices Σ, the comparison is obtained as

$$NIQE(\nu_1, \nu_2, \Sigma_1, \Sigma_2) = \sqrt{(\nu_1 - \nu_2)^{\mathrm{T}} \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1} (\nu_1 - \nu_2)}, \tag{20}$$

where T denotes the operation of vector transposition.

As previously mentioned, the two described naturalness estimators were tested against the overall quality scores for the datasets by Čadík et al.[2] and by Yeganeh and Wang.[13] For comparison also the performance of full-reference TMQI measure and its Structural Similarity (SS) component is included in the analysis. The results are depicted in Tables 3 and 4, the averaged correlation coefficients are shown in Figures 3 and 4.

Several conclusions could be drawn from the analysis. Firstly, NIQE does not seem to be suitable for this application. The statistics incorporated in the estimator do not correspond well with the statistics of tone-mapped images. Secondly, TMQI outperforms the others in most of the cases, which could have been expected, since it has been trained for the purpose. Interesting fact is that the performance of SN for Čadík's dataset is very poor but for Yeganeh's one it sometimes even outperforms TMQI and is comparable to SS.

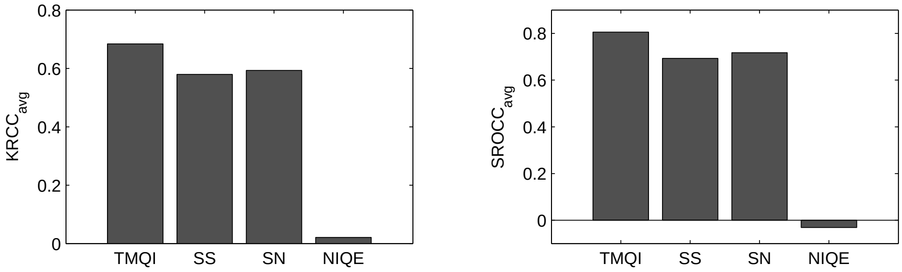
Anyway, the results suggest that more work should be done in the field of reliable tone-mapped images naturalness and quality estimation, possibly by incorporating the above mentioned candidates for features evaluation.

3. CONCLUSION AND FUTURE WORK

This paper discussed the possibilities in no-reference quality assessment of images after tone-mapping. Firstly, an overview of past subjective studies about quality perception and particular features influence in the field was provided. Then, the works concerning evaluation of HDR and tone-mapped images were summarized and the possibility for incorporating several attributes estimators into single overall quality metric has been proposed.

Scene	TMQI		SS		SN		NIQE	
	SROCC	KRCC	SROCC	KRCC	SROCC	KRCC	SROCC	KRCC
1	0.9048	0.7857	0.6667	0.5000	0.9048	0.7857	-0.5000	-0.3571
2	0.7857	0.6429	0.8095	0.7143	0.7619	0.5714	-0.5714	-0.2857
3	0.8095	0.6429	0.2619	0.1429	0.7381	0.5714	-0.0714	-0.0714
4	0.8810	0.7143	0.8571	0.7143	0.8571	0.6429	-0.5714	-0.2857
5	0.7381	0.6429	0.1429	0.1429	0.7381	0.6429	0.4048	0.3571
6	0.9762	0.9286	0.7381	0.6429	0.9524	0.8571	-0.0952	0.0000
7	0.6905	0.5714	0.8810	0.7857	0.6429	0.5000	0.1667	0.2143
8	0.7143	0.5714	0.3333	0.2857	0.7143	0.5714	0.2857	0.2143
9	0.6905	0.5714	0.8571	0.7143	0.3571	0.3571	0.1667	0.1429
10	0.9286	0.8571	0.6667	0.5000	0.9048	0.7857	-0.3333	-0.2857
11	0.8810	0.7143	0.6429	0.4286	0.5476	0.4286	0.5476	0.4286
12	0.7143	0.5714	0.7143	0.5714	0.5714	0.4286	0.7619	0.6429
13	0.6826	0.5455	0.9461	0.8365	0.4311	0.3273	-0.2275	-0.1091
14	0.7381	0.6429	0.9524	0.8571	0.7381	0.6429	-0.1429	-0.0714
15	0.9524	0.8571	0.9286	0.8571	0.9048	0.7857	-0.2857	-0.2143

Table 4. Performance of quality and naturalness measures on Yeganeh's dataset.[13]



(a) Average KRCC (b) Average SROCC

Figure 4. Average correlation coefficients of quality and naturalness measures on Yeganeh's dataset.[13]

Six measures of contrast were tested against the ranks provided by the observers according to the perceived contrast. Most of the measures were not found suitable for the particular purpose. However, Global Contrast Factor[24] measure performed the best from the analyzed and was selected as the top candidate for a possible merge. Another possibly useful estimator could be Root Mean Squared contrast which performs slightly worse but is incomparably computationally less demanding.

Another investigated feature was image's colorfulness. Three colorfulness evaluators were implemented and tested in a similar way. The only measure giving at least decent results was CQE1[19] but there is a lot of space left for improvement in this area.

Lastly, two naturalness metrics were considered. It was shown that the Natural Image Quality Evaluator[31] does not provide the good results in this context and even though the Statistical Naturalness[13] measure is incorporated in the Tone Mapped Image Quality Index[13] designed specifically for this purpose, its performance is limited and the power of the metric lies more in the training.

Next steps will include the fusion of the best possible estimators in one no-reference metric. Also other attributes such as sharpness etc. could be considered. Moreover, it has been shown that the performance of TMQI is good despite the fact it calculates with the luminance component only. The subjective studies however claimed it to be of high importance. This issue should be carefully studied to show if the current datasets do not provide enough variance in colorfulness to test these aspects, or if the information in the luminance component is sufficient for reliable objective quality assessment.

ACKNOWLEDGMENTS

This work was partially supported by the COST IC1003 European Network on Quality of Experience in Multimedia Systems and Services - QUALINET, by the COST CZ LD12018 Modeling and verification of methods for Quality of Experience (QoE) assessment in multimedia systems - MOVERIQ and by the grant No. P102/10/1320 Research and modeling of advanced methods of image quality evaluation of the Czech Science Foundation.

Authors would also like to thank professor Peter Schelkens for his valuable comments and discussion.

REFERENCES

[1] Reinhard, E., Ward, G., Pattanaik, S., Debevec, P., Heidrich, W., and Myszkowski, K., [*High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*], CA: Morgan Kaufmann (2010).

[2] Cadik, M., Wimmer, M., Neumann, L., and Artusi, A., "Evaluation of HDR tone mapping methods using essential perceptual attributes," *Computers & Graphics* **32**, 330–349 (2008).

[3] Drago, F., Martens, W. L., Myszkowski, K., and Seidel, H.-P., "Perceptual evaluation of tone mapping operators," in [*SIGGRAPH conference on sketches & applications*], (2003).

[4] Kuang, J., Yamaguchi, H., Liu, C., Johnson, G. M., and Fairchild, M. D., "Evaluating hdr rendering algorithms," *ACM Transactions on Applied Perception* **4**(2) (2007).

[5] Yoshida, A., Blanz, V., Myszkowski, K., and Seidel, H.-P., "Perceptual evaluation of tone mapping operators with real-world scenes," in [*SPIE Human Vision & Electronic Imaging*], (2005).

[6] Ledda, P., Chalmers, A., Troscianko, T., and Seetzen, H., "Evaluation of tone mapping operators using a high dynamic range display," in [*32nd annual conference on computer graphics and interactive techniques SIGGRAPH*], (2005).

[7] Yoshida, A., Mantiuk, R., Myszkowski, K., and Seidel, H.-P., "Analysis of reproducing real-world appearance on ddisplay of varying dynamic range," *Computer Graphics Forum* **25**(3), 415–426 (2006).

[8] Ashikhmin, M. and Goyal, J., "A reality check for tone-mapping operators," *ACM Transactions on Applied Perception* **3**(4), 399–411 (2006).

[9] Akyüz, A. O., Fleming, R., Riecke, B. E., Reinhard, E., and Bülthoff, H. H., "Do HDR ddisplay support LDR content? a psychophysical evaluation," *ACM Transactions on Applied Perception* **26**(3) (2007).

[10] Barkowsky, M. and Le Callet, P., "On the perceptual similarity of realistic looking tone mapped high dynamic range images," in [*International Conference on Image Processing*], (2010).

[11] Kuang, J., Johnson, G. M., and Fairchild, M. D., "icam06: A refined image appearance model for HDR image rendering," *Journal of Visual Communication and Image Representation* **18**, 406–414 (2007).

[12] Song, M., Tao, D., Chen, C., Bu, J., Luo, J., and Zhang, C., "Probabilistic exposure fusion," *IEEE Transactions on Image Processing* **21**(1), 341–357 (2012).

[13] Yeganeh, H. and Wang, Z., "Objective quality assessment of tone-mapped images," *IEEE Transactions on Image Processing* **22**, 657–667 (February 2013).

[14] Mantiuk, R., Daly, S., Myszkowski, K., and Seidl, H.-P., "Predicting visible differences in high dynamic range images - model and its calibration," in [*Human Vision and Electronic Imaging, SPIE*], (2005).

[15] Daly, S., [*Digital Images and Human Vision*], ch. The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity, 179 – 206, MIT Press (1993).

[16] Aydin, T. O., Mantiuk, R., Myszkowski, K., and Seidl, H.-P., "Dynamic range independent image quality assessment," in [*International Conference on Computer Graphics and Interactive Techniques*], (2008).

[17] Mantiuk, R., Kim, K. J., Rempel, A. G., and Heidrich, W., "Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," in [*ACM Transactions on Graphics (Proc. of SIGGRAPH'11)*], **30**(4) (2011).

[18] Wang, Z., Simoncelli, E., and Bovik, A., "Multi-scale structural similarity for image quality assessment," in [*IEEE Asilomar Conference on Signal, Systems and Computers*], **2**, 1398–1402 (2003).

[19] Panetta, K., Gao, C., and Agaian, S., "No reference color image contrast and quality measures," *IEEE Transactions on Consumer Electronics* **59**, 643 – 651 (August 2013).

[20] Agaian, S., Lentz, K. P., and Grigoryan, A. M., "A new measure of image enhancement," in [*IASTED International Conference on Signal Processing and Communications*], (2000).

[21] Agaian, S., Silver, B., and Panetta, K. A., "Trmeasure coefficient histogram-based image enhancement algorithms using contrast entropy," *IEEE Transactions on Image Processing* **16**, 741 – 758 (2007).

[22] Panetta, K., Zhou, Y., and Agaian, S., "Nonlinear unsharp masking for mammogram enhancement," *IEEE Transactions on Information Technology in Biomedicine* **15**(4), 918 – 928 (2011).

[23] Nercessian, S., Agaian, S. S., and Panetta, K. A., "Multi-scale image enhancement using a second derivative-like measure of contrast," in [*SPIE-IS&T Electronic Imaging: Image Processing: Algorithms and Systems X*], **8295** (2012).

[24] Matković, K., Neumann, L., Neumann, A., Psik, T., and Purgathofer, W., "Global contrast factor - a new approach to image contrast," in [*Proceedings of the First Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging*], *Computational Aesthetics'05*, 159–167, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland (2005).

[25] Hasler, D. and Suesstrunk, S. E., "Measuring colorfulness in natural images," in [*Electronic Imaging*], 87–95 (2003).

[26] Fu, Y.-Y., *Color image quality measures and retrievel*, PhD thesis, Department of Computer Science, New Jersey Institute of Technology (2003).

[27] Srivastava, A., Lee, A. B., Simoncelli, E. P., and Zhu, S.-C., "On advances in statistical modeling of natural images," *J. Math. Imaging and Vision* **18**, 17–33 (January 2003).

[28] Ruderman, D. L., "The statistics of natural images," *Network Computation in Neural Systems* **5**(4), 517 – 548 (1994).

[29] Sheikh, H., Bovik, A., and Veciana, G. D., "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Transactions on Image Processing* **14**, 2117–2128 (December 2005).

[30] Sheikh, H. R., Bovik, A. C., and Cormack, L., "No-reference quality assessment using natural scene statistics: Jpeg2000," *IEEE Transactions on Image Processing* **14**, 1918–1927 (November 2005).

[31] Mittal, A., Soundararajan, R., and Bovik, A. C., "Making a completely blind image quality analyzer," *IEEE Signal Processing Letters* (to appear) (2012).