Gradient-based Image Quality Assessment

Boban Bondzulic[1], Vladimir Petrovic[2], Milenko Andric[1], Boban Pavlovic[1]

[1] Military Academy, University of Defence in Belgrade, Generala Pavla Jurisica Sturma 33, 11000 Belgrade, Serbia

[2] Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia

e-mails: boban.bondzulic@va.mod.gov.rs, v.petrovic@manchester.ac.uk, milenko.andric@va.mod.gov.rs, boban.pavlovic@va.mod.gov.rs

Abstract: An objective measure for image quality assessment based on a direct comparison of visual gradient information in the test and reference images is proposed. A perceptual model is defined to provide local estimates of gradient preservation and investigate perceptual importance pooling of such local quality estimates by using the lowest scores. The proposed perceptual pooled measure is validated using extensive subjective test results. Results indicate that the proposed measure is perceptually meaningful in that it corresponds well with the results of subjective evaluation and can outperform actual objective metrics.

Keywords: gradient magnitude; gradient orientation; image quality assessment

1 Introduction

Recent years have seen tremendous growth in visual information representation and communication applications whose performance depends greatly on the quality of output images. Subjective trials and mean opinion scores (MOS) are the most relevant way of assessing image quality but they are inconvenient, slow, and expensive for most real applications. Objective image quality metrics predict perceived image quality computationally.

Objective image and video quality assessment measures have been used in numerous applications. Most of the applications relate to situations where it is necessary to evaluate the quality of a modified version of the reference (original, source) image or they have been used in situations where a comparison is converted into something that is not the signal quality, such as a set of measured data or decisions [1]. Thus, image and video quality assessment measures have been used in the following applications: steganography, digital image watermarking, image fusion quality assessment, noise removal, image enhancement, the assessment of the success of super-resolution techniques, assessing the quality of the resolution degraded images, dynamic range image conversion, coding, remote sensing, video surveillance, object identification, object tracking, classification, analysis of quality of service, etc.

This paper explores the feasibility of a gradient preservation framework, successfully applied to image fusion evaluation [2], in the domain of objective full-reference image quality assessment. The proposed method is a customized and linearized version of the initial framework and is tested on well known, publicly available subject-rated image databases with different distortion types and levels of distortion [3-7]. The performance is compared with actual image quality assessment measures: peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) [8], as well as its relatives – universal image quality index (UIQI) and multi-scale structural similarity (MS-SSIM) [9], visual information fidelity (VIF) [10], visual signal-to-noise ratio (VSNR) [11], most apparent distortion (MAD) [12] and edge preservation measure (Q^{AB}) [13].

Image gradient has, in recent years, been used in an increasing number of ways in assessing image quality [14-24]. In the largest number of objective measures, the gradient magnitude of the original and test image is evaluated, mostly using Roberts, Sobel, Scharr or Prewitt filters, after which the magnitude comparison is performed in similar manner to the SSIM index [14-16]. In addition to image gradient magnitude, different methods often use additional features. Thus, in [17], in addition to the gradient magnitude, phase congruency was used as a measure of the significance of the local structure and as a complementary feature in the local quality assessment. In [18], gradient magnitude is combined with a visual saliency map, which has a dual role – as a feature to determine local quality of the test image and as a weighting function when pooling local quality scores into a global one. A reliable objective measure from [19], in addition to determining the similarity of the gradient magnitudes, also uses chromaticity similarity to measure color distortions. In [20] gradient magnitude and color similarity maps in contour regions, edge-extension regions and slowly-varying regions are pooled by two complementary aspects: visual saliency and visual masking effect.

Apart from a full reference assessment of test image quality compared to the original signal, gradient magnitude has also been used for reduced-reference [21], and no-reference image quality estimation [22, 23]. The method in [21] exploits natural image statistics and shows that log histogram of natural image gradients obeys a specific distribution. No-reference image quality assessment model from [22] utilizes joint statistics of the normalized gradient magnitude map and the Laplacian of Gaussian response. Another blind image quality assessment approach from [23] extracts features in both the spatial (point-wise statistics) and gradient (neighboring gradient magnitude statistical features) domains. While mostly using complimentary approaches, all these studies agree on the fact that gradient information is key to estimating objective image quality and is particularly useful in comparing structures between original and test images. In this context however only gradient magnitude is used and directional gradient information is ignored.

In this paper an objective, full reference image quality metric based on the preservation of gradient information from the original signal is proposed. Our contribution is to explore and use gradient orientation information as a complementary feature to gradient magnitude as well as new effective methods for pooling of local quality scores obtained using gradient information. The application of gradient orientation has not been fully explored in the context of image quality assessment, with very few studies available in the literature [13, 24]. We show that using gradient orientation can improve the results, increase the correlation with subjective scores of objective image quality assessment based solely on gradient magnitude. We also shown that the correct selection of local quality scores can additionally increase the degree of agreement between subjective and objective quality scores. The performance of the proposed measure is consistent and stable with five publicly available subject-rated image datasets.

2 Theory

Image gradient plays a very important role in human understanding of visual signals, effectively serving to carry structural scene information. As such it is a vital feature in the development of objective quality assessment measures that largely base their measurement on the preservation of this information from the original image into the test image. Different types of degradations lead to a gradient changes, with changes in contrast graded by changes in gradient magnitude, and structural changes evident in changes to gradient orientation. Using estimates of both local gradient magnitude and orientation, local quality of reproduction of the information from the original image can be determined as a direct measure of displayed image quality. In this manned both the contrast and shape distorting effects of various degradations to image quality can be measured.

Gradient preservation framework is based on the idea that only successful transfer of image structures from the reference into the test image constitutes good quality and that structural information can be captured by looking at local intensity gradients. The method extracts gradient information and uses a perceptual change model to compare them in between reference and test images to obtain local estimates of gradient preservation. These effectively local quality estimates are combined using a more advanced perceptual pooling method into an overall objective quality score.

Initially, local x and y gradients are extracted from the reference and test images, R and T, using Sobel templates. Gradient (edge) magnitude, g, and orientation, α, are easily obtained for each pixel (n,m) from the Sobel responses s^x and s^y according to:

$$g_R(n,m)=\frac{\sqrt{s_R^x(n,m)^2+s_R^y(n,m)^2}}{g_{\max}} \tag{1}$$

$$\alpha_R(n,m)=\arctan\left(\frac{s_R^y(n,m)}{s_R^x(n,m)}\right) \tag{2}$$

where $g_{\max}$ is maximum magnitude, taken as $g_{\max}$=4.472, for 8-bits/pixel grayscale images. Both parameters are thus bounded, $g\in[0,1]$ from none to maximum contrast, and orientation $\alpha\in[-\pi, \pi]$.

It is assumed that an input edge is perfectly represented only if both its magnitude and its orientation are unchanged in the test image. When a loss of contrast exists between R into T, gradient magnitude change, Δ_g, is observed, and is defined as:

$$\Delta_g(n,m)=\begin{cases}\dfrac{g_T(n,m)+C}{g_R(n,m)+C}, g_R(n,m)>g_T(n,m)\\[2ex] \dfrac{g_R(n,m)+C}{g_T(n,m)+C}, g_R(n,m)\le g_T(n,m)\end{cases} \tag{3}$$

where the constant C (C=1/64) is included to avoid instability when the denominator in Eq. (3) is very close to zero.

Orientation α however, is cyclic, i.e. values at the two extremes ($-\pi$, and π) are in fact equivalent and change in orientation in T with respect to R, Δ_α, measuring structural similarity can be defined as:

$$\Delta_\alpha(n,m)=\frac{\left|\left|\alpha_R(n,m)-\alpha_T(n,m)\right|-\pi\right|}{\pi} \tag{4}$$

For a total of NxM pixels, the overall success of gradient preservation is obtained as a mean value of local gradient preservations:

$$\Delta_i^{RT}=\frac{1}{NM}\sum\nolimits_{\forall n,m}\Delta_i(n,m),\ i\in\{g,\alpha\} \tag{5}$$

This model in effect quantifies perceived visual information loss with respect to changes in gradient parameters, broadly changes in contrast (magnitude), Δ_g^{RT}, and shape/structure (orientation), Δ_α^{RT}. Gradient magnitude and orientation preservations, Δ_g^{RT} and Δ_α^{RT}, are combined into a single gradient preservation measure Δ^{RT}:

$$\Delta^{RT}=\sqrt{\Delta_g^{RT}\cdot\Delta_\alpha^{RT}} \tag{6}$$

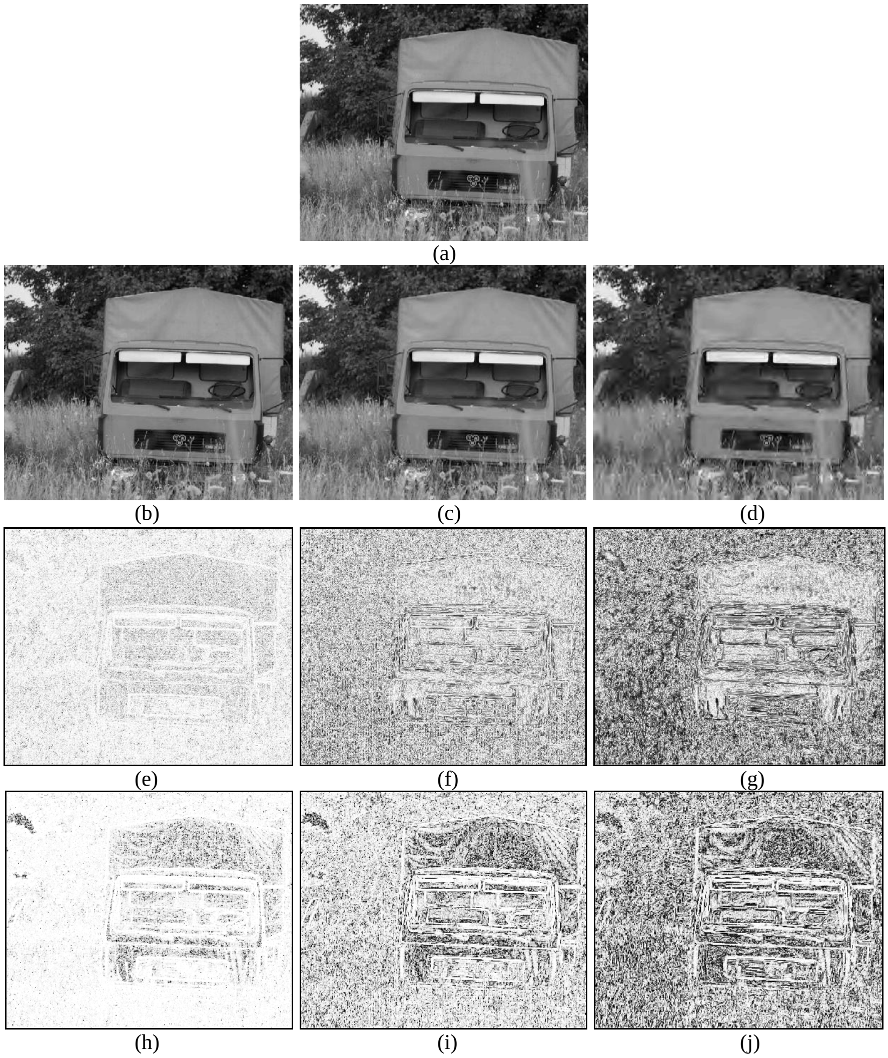

Figure 1

(a) reference image, (b) (c) (d) distorted (test) images (created by JPEG2000 compression), (e) (f) (g) gradient magnitude preservation maps computed using Eq. (3), and (h) (i) (j) gradient orientation preservation maps computed using Eq. (4)

A pixel-domain full-reference example is shown on Figure 1, where the goal is to evaluate the quality of test images, (b), (c) and (d), with a given perfect-quality reference image (a) (images are from the VCL@FER database [5]). The resulting gradient magnitude and orientation information preservation maps are shown below the test images – the brighter indicates better quality (larger local Δ_g^{RT} and Δ_α^{RT} values). The gradient information preservation maps reflect the spatial variations of the perceived image quality. The careful inspection shows that the coarse quantization in JPEG2000 algorithm results in smooth representations of fine-detail regions in the image (e.g. the trees and the grass in (c) and (d)).

Table 1 provides subjective (MOS) and objective values for test images on Fig. 1. Objective measures deliver good consistency with perceived quality measurements. Notice from Table 1 that for high-quality Figure 1(b) image, MAD technique, which uses a simple spatial-domain model of local visual masking, provides the value of 0 (lower is better). It means that there are no visible distortions on Figure 1(b).

Table 1
Subjective (MOS) and objective evaluations for the test images shown on Figure 1

Image	MOS	PSNR	MS-SSIM	VIF	VSNR	MAD	Q^{AB}	Δ_g^{RT}	Δ_α^{RT}	Δ^{RT}
Fig. 1(b)	75.72	43.89	0.99	0.98	42.15	0	0.91	0.95	0.94	0.94
Fig. 1(c)	52.44	31.68	0.98	0.51	33.15	31.86	0.56	0.78	0.79	0.78
Fig. 1(d)	31.59	26.85	0.89	0.18	19.62	69.79	0.31	0.66	0.66	0.66

3 Perceptual Importance Pooling

Image quality assessment is most often carried out in two phases. In the first phase, the quality is determined at the local level, while in the second phase, the integration of local quality scores is performed to determine a single global quality score for the entire test image. The second phase, considered in this chapter focuses on the observation that human observers do not base their impressions of quality on the entire visible signal. Furthermore, the influence different locations in the signal have on their subjective scores varies highly [25, 26] and in order to predict subjective quality scores, this effects needs to be modeled effectively. In addition to the most obvious average pooling of all local quality scores, different techniques for the association of local quality scores have been proposed: deviation based pooling, region-based pooling, pooling using the lowest quality scores, ... [26].

Summations in Eq. (5), effectively represent a linear spatial pooling where each pixel has an equal influence on the overall quality score. It is an established fact however, that humans tend to attach more importance to regions of poor quality in images [25, 26]. Perceptual importance approach by pooling over only the lowest Δ_g and Δ_α scores, i.e. only regions with poor quality, is investigated. Specifically, quality maps Δ_g and Δ_α are found using Equations (3) and (4), then the values are arranged in ascending order. A mean score is calculated from the lowest p% of these values ($\Delta_g^{p\%}/\Delta_\alpha^{p\%}$). Pixels that fall outside this percentile range are rejected.

Driven by the experience of probabilistic systems where a single low value biases a global score obtained using a product rule (e.g. Eq. (6)), a simpler, additive framework as an alternative to Eq. (6), combined with optimal quality guided lowest percentile pooling is investigated:

$$AM\text{-}\Delta^{RT} = w_g \Delta_g^{p_g\%} + (1 - w_g)\Delta_\alpha^{p_\alpha\%} \tag{7}$$

where, w_g and (1-w_g) are the relative importance of the magnitude and orientation components, $w_g \in [0,1]$. Two questions remain – what percentile should be used and what weight to assign to each of the two components?

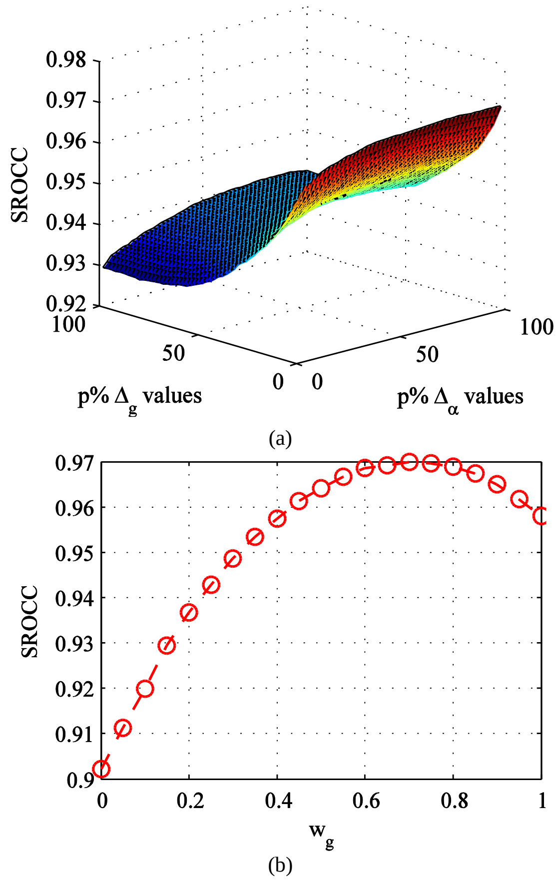

Figure 2

(a) SROCC as a function of the lowest p% scores for Δ_g and Δ_α (in p=2% increments and for optimal w_g=0.7 value), and (b) training set SROCC of AM-Δ^{RT} (Eq. (7)) as function of w_g (in 0.05 increments and for optimal p_g=2% and p_α=78% values)

In order to determine the optimal values for p_g, p_α and w_g, an exhaustive optimization on LIVE image quality assessment database [3] was performed. Fifteen reference images and their distorted versions were selected for training (374 images) and parameters that produce optimal Spearman's rank-order correlation coefficient (SROCC) for the proposed AM metric were sought. Optimum values that were obtained are p_g=2%, p_α=78%, and w_g=0.7. A p_g–p_α section of the 3D optimization surface for $AM\text{-}\Delta^{RT}$ at w_g=0.7, is illustrated on Figure 2(a). Low p_g values provide the most relevant quality measurements while the robust performance is observed over the entire p_α range. The effect of weight distribution between the Δ_g and Δ_α channels is illustrated on Figure 2(b) showing SROCC for $AM\text{-}\Delta^{RT}$ as function of w_g (w_α=1-w_g) at p_g=2% and p_α=78% values. Optimally, contrast measure Δ_g is marginally more important than local structure Δ_α, 0.7 vs. 0.3.

4 Results

To demonstrate the performance of the proposed measure, the rest of LIVE image quality assessment database [3] – fourteen reference images and their distorted versions for testing (405 images), CSIQ [4], VCL@FER [5], MCL-JCI [6] and JPEG XR [7] image quality assessment databases were used.

Table 2 provides the comparison of the used, publicly available databases. Databases have different numbers of reference images (6-50), distorted (rated) images (180-866), distortion types (1-6), distortion levels (3-9), number of human observers, ratings and stimulus method. Viewing conditions (e.g. display resolution and viewing distance) are different also.

Subjective tests where average human observers are displayed series of test, and optionally corresponding original images and their quality impressions of those images collected as simple scalar ratings have long been considered as the most reliable way to obtain ground truth evaluation of perceptual image quality. Individual subjective quality scores (opinions) are usually sumarised in the form of mean opinion values of the scores, MOS/DMOS/SQF, and confidence intervals about those scores for each evaluated image. Subjective trials are usually conducted in strictly controlled environmental conditions and involve large user samples to render statistical relevance to their results, making them time and effort consuming and impractical for any routine use in imaging applications. The goal of objective metrics has always been accurate prediction of such scores that could be obtained without the complex practical procedure involved in organizing subjective trials. Subjective studies conducted so far have mostly been inconclusive in terms of identifying a single optimal objective metric [4-7] with various metrics exhibiting optimal performance for different sets of subjectively evaluated data.

Table 2
Comparison of the public databases

	LIVE	**CSIQ**	**VCL@FER**	**MCL-JCI**	**JPEG XR**
Year	2006	2009	2011	2016	2009
Display	CRT, 21"	Sceptre 24", X24WG LCD	N/A	65"	Eizo CG301W LCD
Display resolution	1024x768	1920x1200	N/A	3840x2160	2560x1600
Viewing distance	2-2.5 SH[!]	70 cm	N/A	2 m (~1.6 SH)	1 SH
Reference images	29	30	23	50	10 (4 for training and 6 for testing)
Image resolution	~768x512	512x512	~768x512	1920x1080	1600x1280
Distortion types	JPEG, JPEG2000, additive Gaussian noise, blurring, fast fading	JPEG, JPEG2000, blurring, contrast decrements, additive pink noise, additive Gaussian noise	JPEG, JPEG2000, blurring, additive Gaussian noise	JPEG	JPEG, JPEG2000 (two configurations), JPEG XR (two implementations)
Distortion levels	5-9	4-5	6	3-7	6
Method	Single Stimulus (with hidden reference)	Categorical Subjective Image Quality	Single Stimulus	two images (side by side)	Double-Stimulus Continuous Quality Scale
Data	DMOS[%]	DMOS	MOS	SQF[&]	MOS
Observers	161	35	118	>150	16
Number of ratings per image	20-29	5-7	16-36	30	16
Test images	779	866	552	243	180
Format	**BMP**	**PNG**	**BMP/JPG**	**BMP**	**BMP**

N/A = Not Available   [!] SH = Screen Height

[%] DMOS = Difference MOS   [&] SQF = Stair Quality Function [6]

The performance of objective metrics was evaluated over three aspects of their ability to estimate subjective image quality [27]: (i) prediction accuracy, measured using linear correlation coefficient (LCC), mean absolute error (MAE), and root mean squared error (RMSE); (ii) prediction monotonicity, measured using the SROCC; and (iii) prediction consistency, quantified using the outlier ratio (OR).

A comparison over five performance measures of several objective metrics on LIVE test images is summarized in Table 3 (three best methods are in bold). *AM-*Δ^{RT} outperforms other objective measures. In contrast to some prior studies [28], significant gains in performance can be obtained using the right pooling strategy, compare the Δ^{RT} and AM-Δ^{RT} scores. The significance in using both gradient magnitude and orientation information can be seen in the difference between complete metrics Δ^{RT} and AM-Δ^{RT} on one side and Δ_g^{RT} and Δ_α^{RT} on the other.

Table 3

Performance comparison on LIVE test images (405 images) [3]

Method	LCC	SROCC	MAE	RMSE	OR (%)
PSNR	0.8784	0.8852	10.1182	13.0942	9.8765
UIQI	0.8982	0.8925	9.3335	12.0433	6.9136
SSIM	0.9008	0.9107	9.2729	11.8967	7.6543
MS-SSIM	0.9443	0.9596	7.2996	9.0185	2.7160
VIF	**0.9623**	**0.9662**	**6.0976**	**7.4502**	**0.2469**
VSNR	0.9265	0.9320	7.9889	10.3109	4.1975
MAD	**0.9648**	**0.9652**	**5.5983**	**7.2002**	**0.4938**
Q^{AB}	0.9405	0.9418	7.5332	9.3083	2.9630
Δ_g^{RT}	0.9190	0.9235	8.3077	10.8030	4.6914
Δ_α^{RT}	0.9235	0.9150	8.4850	10.5109	3.2099
Δ^{RT}	0.9403	0.9443	7.3584	9.3216	2.7160
AM-Δ^{RT}	**0.9692**	**0.9709**	**5.4455**	**6.7419**	**0.2469**

Table 4

Performance comparison on CSIQ images [4]

Method	LCC	SROCC	MAE	RMSE	OR (%)
PSNR	0.7999	0.8057	0.1195	0.1576	34.2956
UIQI	0.8289	0.8092	0.1127	0.1469	34.4111
SSIM	0.8151	0.8368	0.1161	0.1521	33.4873
MS-SSIM	0.8666	**0.8774**	**0.0972**	0.1310	**27.7136**
VIF	**0.9252**	**0.9194**	**0.0753**	**0.0996**	**22.7483**
VSNR	0.8018	0.8132	0.1152	0.1569	30.1386
MAD	**0.9502**	**0.9466**	**0.0636**	**0.0818**	**17.8984**
Q^{AB}	0.8556	0.8520	0.1039	0.1359	31.1778
Δ_g^{RT}	0.8459	0.8690	0.1052	0.1400	30.1386
Δ_α^{RT}	0.7792	0.7147	0.1332	0.1646	40.9931
Δ^{RT}	0.8605	0.8621	0.1018	0.1338	29.4457
AM-Δ^{RT}	**0.8847**	0.8616	0.0986	**0.1224**	29.9076

Tables 4–7 provide further objective metric performance results on CSIQ [4], VCL@FER [5], MCL-JCI [6], and JPEG XR [7] databases (AM-Δ^{RT} uses parameters determined on the LIVE training set). Combined magnitude/orientation models achieve better results than individual preservation models (Δ_α^{RT} and Δ_g^{RT}). The additive combined model, Eq. (7), outperforms the

multiplicative, Eq. (6), and achieves performance near the top of the tested metrics (MS-SSIM, VIF and MAD).

Table 5
Performance comparison on VCL@FER images [5]

Method	LCC	SROCC	MAE	RMSE	OR (%)
PSNR	0.8321	0.8246	10.2335	13.6204	53.8043
UIQI	0.7965	0.7983	11.5681	14.8495	62.5000
SSIM	0.8742	0.8677	9.3849	11.9244	54.8913
MS-SSIM	**0.9183**	**0.9227**	**7.7862**	**9.7238**	**49.0942**
VIF	0.8922	0.8866	8.8811	11.0905	53.9855
VSNR	0.8805	0.8754	8.9194	11.6415	**52.1739**
MAD	**0.9051**	**0.9061**	**8.2371**	**10.4450**	**49.6377**
$\mathbf{Q^{AB}}$	0.8694	0.8692	9.6409	12.1358	59.9638
$\mathbf{\Delta_g^{RT}}$	0.8819	0.8723	9.0247	11.5790	53.8043
$\mathbf{\Delta_\alpha^{RT}}$	0.8055	0.8039	11.2442	14.5545	61.0507
$\mathbf{\Delta^{RT}}$	0.8898	0.8879	8.9453	11.2091	56.1594
AM-$\mathbf{\Delta^{RT}}$	**0.9036**	**0.8978**	**8.3128**	**10.5201**	52.3551

Table 6
Performance comparison on MCL-JCI images [6]

Method	LCC	SROCC	MAE	RMSE
PSNR	0.4721	0.4486	0.1907	0.2288
UIQI	0.5746	0.5713	0.1742	0.2124
SSIM	0.6053	0.5898	0.1676	0.2066
MS-SSIM	0.8340	0.8139	0.1102	0.1432
VIF	**0.8884**	**0.8791**	**0.0909**	**0.1191**
VSNR	0.6441	0.6337	0.1608	0.1985
MAD	**0.8713**	**0.8668**	**0.0984**	**0.1274**
$\mathbf{Q^{AB}}$	0.7879	0.7863	0.1223	0.1598
$\mathbf{\Delta_g^{RT}}$	0.8246	0.7959	0.1138	0.1468
$\mathbf{\Delta_\alpha^{RT}}$	0.6567	0.6466	0.1551	0.1957
$\mathbf{\Delta^{RT}}$	0.8318	0.8229	0.1091	0.1440
AM-$\mathbf{\Delta^{RT}}$	**0.8603**	**0.8462**	**0.1020**	**0.1323**

Table 7
Performance comparison on JPEG XR images [7]

Method	LCC	SROCC	MAE	RMSE	OR (%)
PSNR	0.7819	0.7980	12.8737	16.5360	35.5556
UIQI	0.8621	0.8186	9.5605	13.4404	23.3333

SSIM	0.8744	0.8435	9.6144	12.8684	23.8889
MS-SSIM	0.9309	0.8930	7.0745	9.6863	14.4444
VIF	**0.9389**	**0.9130**	**6.8067**	**9.1278**	13.3333
VSNR	0.8765	0.7803	10.1065	12.7692	23.3333
MAD	**0.9466**	**0.9406**	**6.2598**	**8.5498**	**11.1111**
$\mathbf{Q^{AB}}$	0.9269	0.8995	6.8809	9.9561	**11.6667**
$\mathbf{\Delta_g^{RT}}$	0.9246	0.9071	7.4744	10.1074	12.7778
$\mathbf{\Delta_\alpha^{RT}}$	0.9034	0.8685	7.8474	11.3751	16.1111
$\mathbf{\Delta^{RT}}$	**0.9339**	**0.9117**	**6.5091**	**9.4860**	**9.4444**
AM-$\mathbf{\Delta^{RT}}$	0.9277	0.9089	7.3250	9.9039	12.7778

It is worth noting that no single metric performs best on all the datasets, which is an indication of the sensitivity of the metrics to test data content. The proposed gradient preservation metric with alternative quality guided pooling method *AM-*Δ^{RT} exhibits consistently high performance. Except for the LIVE dataset, gradient magnitude preservation model Δ_g^{RT} provides significantly better results than gradient orientation preservation model Δ_α^{RT}. Hence, it is expected that with improvements of the orientation comparison model, proposed method will improve too.

Furthermore, *AM-*Δ^{RT} is a very well behaved metric with a smooth relationship between objective and subjective scores across the entire range, as shown on the scatter plots on Figure 3.

Since all databases contain JPEG distortion, the performance of objective quality metrics on the JPEG subsets of the five databases was analyzed in more detail. Figure 4 presents subjective-objective agreement (LCC and SROCC) for the eight objective measures on the JPEG subsets (LIVE – 92 images, CSIQ – 150 images, VCL@FER – 138 images, MCL-JCI – 243 images, and JPEG XR – 30 images).

As expected from previous research [29], the performance of quality metrics exhibits similar behavior for the five publicly available databases (extended sets of objective quality measures, databases, and images here were analyzed). The differences over databases, particularly the decrease of performance on MCL-JCI for all objective measures might be explained by a new methodology for perceptual quality measurement – subjective results are given through the stair quality functions (SQF), which are obtained by analysis and post-processing of the raw just noticeable difference (JND) data [6, 30]. Additionally, MCL-JCI dataset contains images with higher spatial resolution than standard datasets used in image quality assessment (see Table 2).

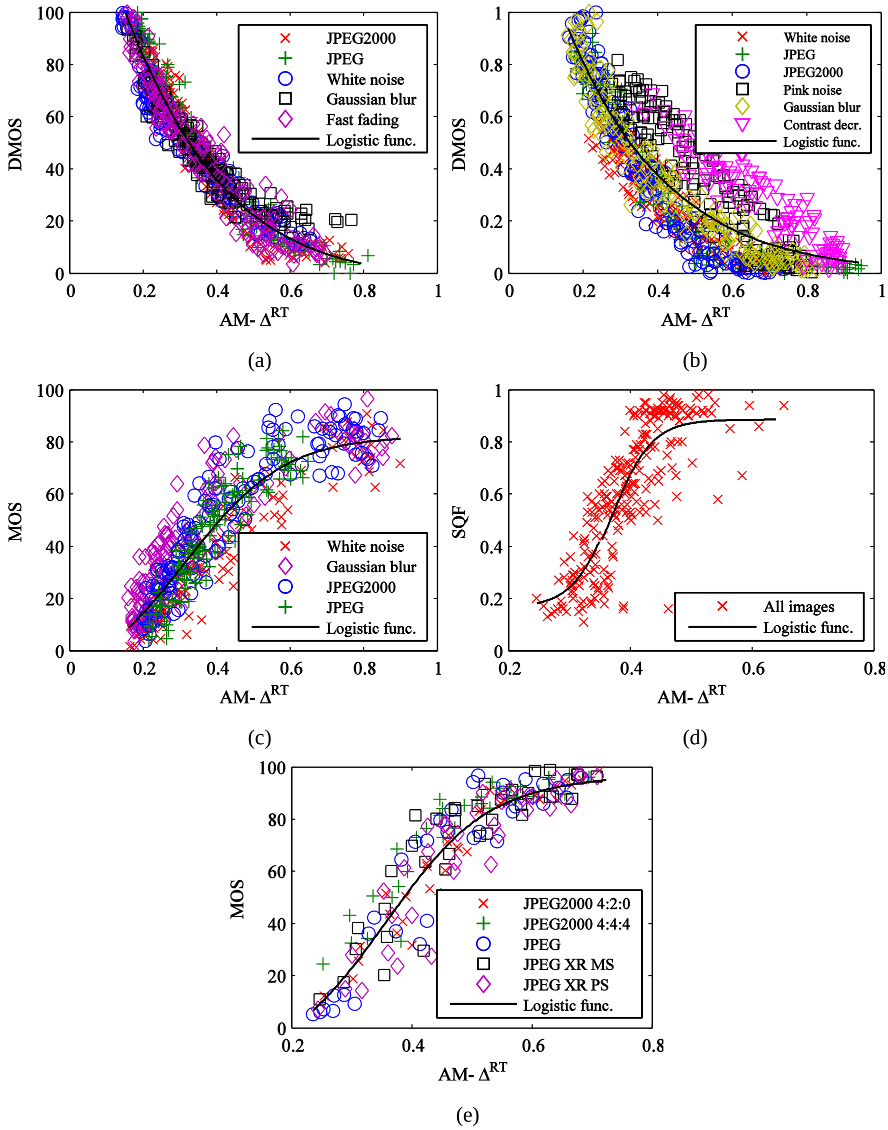

Figure 3

Subjective (DMOS/MOS/SQF) scores versus $AM\text{-}\Delta^{RT}$ model predictions for data from: (a) LIVE, (b) CSIQ, (c) VCL@FER, (d) MCL-JCI and (e) JPEG XR image databases

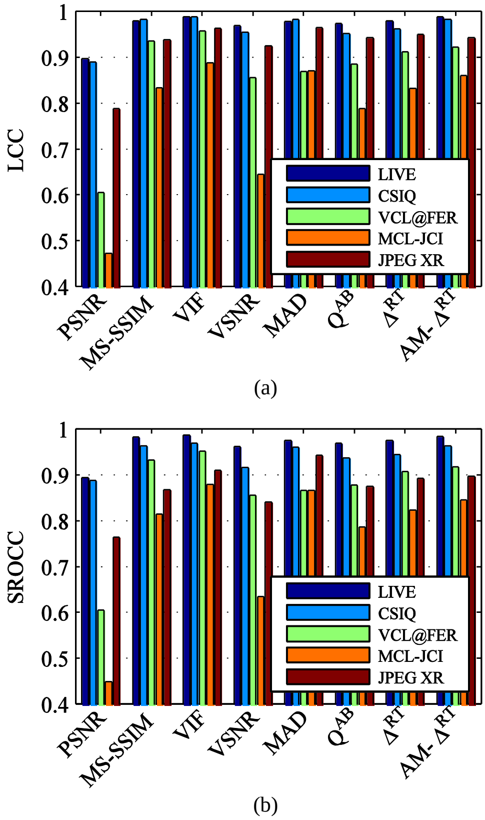

Figure 4

Subjective-objective agreement on the JPEG subsets of the five databases: (a) linear correlation coefficient (LCC), and (b) Spearman's rank-order correlation coefficient (SROCC)

Conclusions

This paper described a novel, gradient-based, full-reference image quality assessment measure, explicitly incorporating gradient orientation information from test signals. Different gradient formulations were investigated, as well as, different spatial score pooling strategies on a variety of subjectively evaluated datasets.

The addition of the gradient orientation information, as a complementary feature to gradient magnitude, is shown to directly improve objective metric performance. Improvement is obtained for all datasets tested in the range 1–3% which is particularly significant in the critical top 15% to the theoretical maximum of the linear correlation range (>0.85).

In contrast to prior studies, it was found that perceptual importance pooling strategy can further improve metric correlation with subjective judgment in a range typically ~3% of linear and rank correlation. Experimental results show that the proposed method achieves consistently high levels of performance, with correlation levels up to 97% and above 85% on all datasets, outperforming many similarly complex metrics and reaching the level of much more complex metric formulations such as VIF and MAD.

Finally, we confirmed a significant variability of metric performance levels on different subjective databases. Significant performance level differences were confirmed to exist in JPEG image subsets. This leads to the conclusion, that metric evaluation on a single subject-rated database is generally insufficient.

In future work, the existing gradient-based assessment approach will be expanded to explore and include explicit formulations for temporal gradient, with the aim of evaluating quality of dynamic, video signals. These studies will also include a critical comparison of the types of gradient assessment models required for static and dynamic imagery.

Acknowledgement

This research has been a part of the project No. VA-TT/1-17-19 supported by the Ministry of Defence, Republic of Serbia.

References

[1] Wang, Z., Bovik, A. C.: Mean squared error: love it or leave it? A new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 2009, Vol. 26, No. 1, pp. 98-117. DOI: 10.1109/MSP.2008.930649

[2] Petrovic, V. S., Xydeas, C.: Objective evaluation of signal-level image fusion performance. *Optical Engineering*, 2005, Vol. 44, No. 8, pp. 087003-(1-8). DOI: 10.1117/1.2009764

[3] Sheikh, H. R., Wang, Z., Cormack, L., Bovik, A. C.: *LIVE image quality assessment database release 2*. [Online] Cited 2010-11-17. Available at: http://live.ece.utexas.edu/research/

[4] Larson, E. C., Chandler, D. M.: *The CSIQ image database*. [Online] Cited 2015-11-17. Available at: http://vision.okstate.edu/?loc=csiq

[5] Zaric, A., Tatalovic, N., Brajkovic, N., et al.: VCL@FER image quality assessment database. *Automatica*, 2012, Vol. 53, No. 4, pp. 344-354, DOI: 10.7305/automatika.53-4.241

[6] Jin, L., Lin, J. Y., Hu, S., et al.: Statistical study on perceived JPEG image quality via MCL-JCI dataset construction and analysis. In *Proceedings of the IS&T International Symposium on Electronic Imaging – Image Quality and System Performance XIII*. San Francisco (CA, USA) 2016, IQSP-222.1-IQSP-222.9

[7] De Simone, F., Goldmann, L., Baroncini, V., Ebrahimi, T.: *JPEG core experiment for the evaluation of JPEG XR image coding*. [Online] Cited 2016-06-17. Available at: http://mmspg.epfl.ch/iqa

[8] Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P.: Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004, Vol. 13, No. 4, pp. 600-612, DOI: 10.1109/TIP.2003.819861

[9] Wang, Z., Simoncelli, E. P., Bovik, A. C.: Multi-scale structural similarity for image quality assessment. In *Proceedings of the 37th Asilomar Conference on Signals, Systems and Computers*. Pacific Grove (CA, USA), 2003, pp. 1398-1402, DOI: 10.1109/ACSSC.2003.1292216

[10] Sheikh, H. R., Bovik, A. C.: Image information and visual quality. *IEEE Transactions on Image Processing*, 2006, Vol. 15, No. 2, pp. 430-444, DOI: 10.1109/TIP.2005.859378

[11] Chandler, D. M., Hemami, S. S.: Vsnr: a wavelet-based visual signal-to-noise ratio for natural images. *IEEE Transactions on Image Processing*, 2007, Vol. 16, No. 9, pp. 2284-2298, DOI: 10.1109/TIP.2007.901820

[12] Larson, E. C., Chandler, D. M.: Most apparent distortion: full reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, 2010, Vol. 19, No. 1, pp. 011006-1-011006-21

[13] Bondzulic, B., Petrovic, V.: Edge-based objective evaluation of image quality. In *Proceedings of the IEEE International Conference on Image Processing*. Brussels (Belgium) 2011, pp. 3305-3308, DOI: 10.1109/ICIP.2011.6116378

[14] Liu, A., Lin, W., Narwaria, M.: Image quality assessment based on gradient similarity. *IEEE Transactions on Image Processing*, 2012, Vol. 21, No. 4, pp. 1500-1512, DOI: 10.1109/TIP.2011.2175935

[15] Zhang, X., Feng, X., Wang, W., Xue, W.: Edge strength similarity for image quality assessment. *IEEE Signal Processing Letters*, 2013, Vol. 20, No. 4, pp. 319-322, DOI: 10.1109/LSP.2013.2244081

[16] Xue, W., Zhang, L., Mou, X., Bovik, A.C.: Gradient magnitude similarity deviation: a highly efficient perceptual image quality index. *IEEE Transactions on Image Processing*, 2014, Vol. 23, No. 2, pp. 684-695, DOI: 10.1109/TIP.2013.2293423

[17] Zhang, L., Zhang, L., Mou, X., Zhang, D.: FSIM: a feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 2011, Vol. 20, No. 8, pp. 2378-2386, DOI: 10.1109/TIP.2011.2109730

[18] Zhang, L., Shen, Y., Li, H.: VSI: a visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 2014, Vol. 23, No. 10, pp. 4270-4281, DOI: 10.1109/TIP.2014.2346028

[19] Nafchi, H. Z., Shankolaei, A., Hedjam, R., Cheriet, M.: Mean deviation similarity index: efficient and reliable full-reference image quality

evaluator. *IEEE Access*, 2016, Vol. 4, pp. 5579-5590, DOI: 10.1109/ACCESS.2016.2604042

[20] Shi, Z., Zhang, J., Cao, Q., Pang, K., Luo, T.: Full-reference image quality assessment based on image segmentation with edge feature. *Signal Processing*, 2018, Vol. 145, pp. 99-105, DOI: 10.1016/j.sigpro.2017.11.015

[21] Cheng, G., Huang, J., Liu, Z., Lizhi, C.: Image quality assessment using natural image statistics in gradient domain. *AEU – International Journal of Electronics and Communications*, 2011, Vol. 65, No. 5, pp. 392-397. DOI: 10.1016/j.aeue.2010.05.007

[22] Xue, W., Mou, X., Zhang, L., Bovik, A. C.: Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features. *IEEE Transactions on Image Processing*, 2014, Vol. 23, No. 11, pp. 4850-4862, DOI: 10.1109/TIP.2014.2355716

[23] Jia, H., Sun, Q., Ji, Z., Wang, T., Chen, Q.: No-reference image quality assessment based on natural scene statistics and gradient magnitude similarity. *Optical Engineering*, 2014, Vol. 53, No. 11, pp. 113110-(1-9), DOI: 10.1117/1.OE.53.11.113110

[24] Liu, L., Hua, Y., Zhao, Q., Huang, H., Bovik, A. C.: Blind image quality assessment by relative gradient statistics and adaboosting neural network. *Signal Processing: Image Communication*, 2016, Vol. 40, pp. 1-15, DOI: 10.1016/j.image.2015.10.005

[25] Moorthy, A. K., Bovik, A. C.: Visual importance pooling for image quality assessment. *IEEE Journal on Selected Topics in Signal Processing*, 2009, Vol. 3, No. 2, pp. 193-201, DOI: 10.1109/JSTSP.2009.2015374

[26] Bondzulic, B., Petrovic, V.: Additive models and separable pooling, a new look at structural similarity. *Signal Processing*, 2014, Vol. 97, No. 4, pp. 110-116, DOI: 10.1016/j.sigpro.2013.10.020

[27] ITU-T Telecommunication Standardization Bureau: *Objective Perceptual Assessment of Video Quality: Full Reference Television*. Geneva, Switzerland, 2004

[28] Wang, Z., Shang, X.: Spatial pooling strategies for perceptual image quality assessment. In *Proceedings of the IEEE International Conference on Image Processing*. Atlanta (GA, USA) 2006, pp. 2945-2948 DOI: 10.1109/ICIP.2006.313136

[29] Tourancheau, S., Autrusseau, F., Parvez Sazzad, Z. M., Horita, Y.: Impact of subjective dataset on the performance of image quality metrics. In *Proceedings of the IEEE International Conference on Image Processing*. San Diego (CA, USA) 2008, pp. 365-368, DOI: 10.1109/ICIP.2008.4711767

[30] Lin, J. Y., Jin, L., Hu, S., et al.: Experimental design and analysis of JND test on coded image/video. In *Proceedings of SPIE 9599, Applications of Digital Image Processing XXXVIII*. San Diego (CA, USA) 2015, pp. 95990Z. DOI: 10.1117/12.2188389