

# REGIONAL ATTENTION TO STRUCTURAL DEGRADATIONS FOR PERCEPTUAL IMAGE QUALITY METRIC DESIGN

Ulrich Engelke<sup>†</sup>, Vuong Xuan Nguyen<sup>\*</sup>, and Hans-Jürgen Zepernick<sup>†</sup>

<sup>†</sup>Blekinge Institute of Technology, PO Box 520, 372 25 Ronneby, Sweden, E-mail: uen@bth.se

<sup>\*</sup>University of Duisburg-Essen, Forsthausweg 2, 47057 Duisburg, Germany

## ABSTRACT

In this paper, regional attention to structural degradations in images is analyzed to improve perceptual quality prediction performance of objective image quality metrics. Subjective experiments were conducted to identify regions-of-interest for a set of natural images. A region-selective metric design is then applied to four objective image quality metrics which were trained and validated with respect to quality prediction accuracy and generalization to unknown images. For this purpose, data is used from subjective quality experiments conducted at two independent laboratories. It is shown that the region-selective design is highly beneficial for the considered objective image quality metrics, in particular, prediction accuracy can be significantly increased.

**Index Terms**— Objective image quality metrics, region-of-interest, subjective experiments, feature extraction.

## 1. INTRODUCTION

In natural images, objects that attract peoples attention are commonly referred to as region-of-interest (ROI). This attraction is due to many influencing factors of which some of the strongest are contrast, shape, size, and location of the object. In particular, humans and their faces, have been shown to strongly draw the viewers attention [1]. This phenomenon has extensively been utilized for ROI image coding where the ROI receives a higher coding bit rate than the background (BG) which is particularly useful for image communication when bandwidth is scarce. In order to evaluate the gain through ROI coding, one needs appropriate metrics that are able to independently assess the quality in ROI and BG. Objective image quality metrics, however, are mostly designed to perform the quality prediction on the whole image. This does not agree well with the properties of the human visual system (HVS) which is highly space variant in sampling and processing of visual signals. In fact, the spatial acuity is highest around the central fixation point, the fovea, and decreases strongly with increasing eccentricity [2]. This indicates that image artifacts in the ROI may be perceived more severe than artifacts outside the ROI. Furthermore, it has been shown that the HVS is well adapted to extraction of structural information [3].

Considering the above, the aim of this paper is to determine the impact of structural degradations on perceptual image quality in ROI and BG to enable region-selective image quality metric design. For this purpose, a subjective experiment has been performed to identify ROI for a set of natural images. Four objective image quality metrics are used for the region-selective metric design. The metrics are trained and the quality prediction performance is validated with data from two independent subjective quality experiments. We observed that the region-selective quality metric design substantially increases quality prediction performance of the metrics.

The paper is organized as follows. Section 2 discusses subjective experiments for image quality and ROI identification. Section 3 introduces region-selective objective image quality. In Section 4, quality metric design and prediction performance are discussed. Section 5 concludes the paper.

## 2. SUBJECTIVE EXPERIMENTS

### 2.1. Subjective image quality experiments

The design of objective quality metrics presented in this paper is supported using mean opinion scores (MOS) obtained in subjective quality experiments from two independent laboratories. The first experiment was conducted at Blekinge Institute of Technology (BIT) in Ronneby, Sweden, and the other at the Western Australian Telecommunications Research Institute (WATRI) in Perth, Australia [4]. Each experiment involved 30 non-expert viewers. The experiment procedures were designed according to ITU-R Rec. BT.500-11 [5]. A set  $\mathcal{I}_R$  of 7 reference monochrome images of dimensions  $512 \times 512$  pixels was chosen to account for different textures and complexities. The images were encoded into Joint Photographic Experts Group (JPEG) format. A simulation model of a wireless system was used to generate two sets  $\mathcal{I}_B$  and  $\mathcal{I}_W$  of 40 distorted images each, for BIT and WATRI experiments, respectively. In particular, blocking, blur, ringing, and intensity masking artifacts were observed in different degrees of severity. The viewers were shown the distorted images along with their reference images. The experiments at BIT and WATRI resulted in two respective sets of MOS,  $\mathcal{M}_B$  and  $\mathcal{M}_W$ .



Fig. 1. Mean ROI for the images in  $\mathcal{I}_R$  (black frame: before outlier elimination; brightened area: after outlier elimination).

Table 1. Statistical analysis of ROI experiment.

	$\mu_{x_C}$	$\sigma_{x_C}$	$\mu_{y_C}$	$\sigma_{y_C}$	$r_0$
Barbara	350	99.92	344	89.51	16.7
Elaine	260	42.24	263	46.96	10
Goldhill	288	65.36	204	87.79	10
Lena	278	60.2	227	31.99	10
Mandrill	256	8.27	339	86.14	3.3
Pepper	235	84.15	262	58.85	10
Tiffany	316	33.1	231	52.95	3.3

## 2.2. Subjective experiment for ROI identification

A subjective ROI experiment was conducted at BIT where viewers had to select an image region that draws their attention. The outcomes enabled us to identify a rectangular mean ROI for each of the reference images in  $\mathcal{I}_R$  and ultimately to perform the region-selective metric design. The experiment involved 30 non-expert viewers and comprised of three trials; training, stabilization, and test. A simple training image was used to explain the ROI selection process, followed by two stabilization images for the viewer to adapt to the process. The actual test set comprised of the reference images in  $\mathcal{I}_R$ .

For each of the images it was observed that a few selections were far away from the majority of the votes. These selections, also referred to as outliers, were eliminated by adopting the criterium defined in [6] as follows

$$|x_C - \mu_{x_C}| > 2 \cdot \sigma_{x_C} \quad \text{or} \quad |y_C - \mu_{y_C}| > 2 \cdot \sigma_{y_C} \quad (1)$$

where  $x_C$  and  $y_C$  are ROI center point coordinates in horizontal and vertical direction, respectively, with the origin in the bottom left image corner. Furthermore,  $\mu$  and  $\sigma$  denote the corresponding mean and standard deviation over all 30 ROI selections, respectively. Based on the number of eliminated outliers we define an outlier ratio for each of the images as

$$r_0 = \frac{N_0}{N} \cdot 100 \quad [\%] \quad (2)$$

where  $N_0$  is the number of eliminated ROI selections and  $N$  the number of all ROI selections. A statistical analysis of the experiment is summarized in Table 1. The mean ROI are shown in Fig. 1 where the black frame and brightened region emphasize the mean ROI before and after outlier elimination,

respectively. It should be noted, that in order for the objective quality metrics (see Section 3.1) to produce meaningful results, ROI were adjusted to fall into the closest  $8 \times 8$  block borders produced by the discrete cosine transform of the JPEG coder. However, considering the image size, the maximum error due to this necessary adjustment is only 0.78%.

## 3. REGION-SELECTIVE OBJECTIVE QUALITY

### 3.1. Objective image quality metrics

For the region-selective metric design we considered the following four objective image quality metrics which were originally designed for quality assessment of a whole distorted image  $I_D$  as compared to a whole reference image  $I_R$ .

**Metric 1:** The Normalized Hybrid Image Quality Metric (NHIQM) has been proposed in [7]. It is based on extraction of five structural features  $f_n \in [0, 1]$ , in particular, blocking, blur, edge-based image activity, gradient-based image activity, and intensity masking. The individual feature measures are normalized and accumulated resulting in a single value

$$NHIQM = \sum_{n=1}^5 w_n \cdot f_n \quad (3)$$

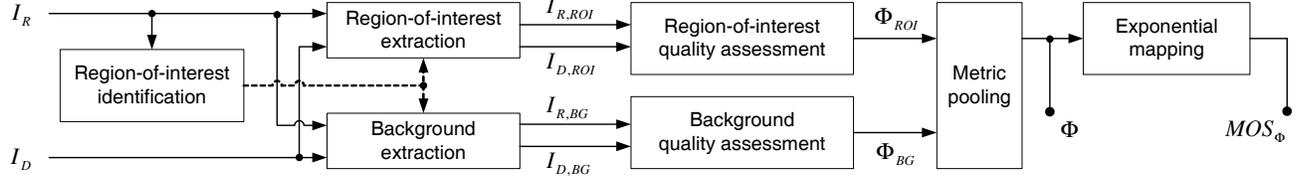
where the weights  $w_n$  regulate the impact of a feature on the overall metric. More precisely, weights  $w_n$  were derived as Pearson linear correlations of the corresponding features  $f_n$  with MOS  $\mathcal{M}_B$ , as a means for the perceptual relevance of a feature [7]. Further defined is an absolute difference as a measure of structural degradations between two images

$$\Delta_{NHIQM} = |NHIQM_R - NHIQM_D|. \quad (4)$$

**Metric 2:** A reduced-reference image quality assessment (RRIQA) technique is described in [8] which is based on a natural image statistic model in the wavelet domain. The distortion between two images is calculated as

$$RRIQA = \log_2 \left( 1 + \frac{1}{D_0} \sum_{k=1}^K |\hat{d}^k(p^k \| q^k)| \right) \quad (5)$$

where the constant  $D_0$  is a scaler of the distortion measure,  $K$  is the number of subbands, and  $\hat{d}^k(p^k \| q^k)$  an estimation of the Kullback-Leibler distance between the probability density functions  $p^k$  and  $q^k$  of the  $k^{th}$  subband in the two images.



**Fig. 2.** Overview of the image quality assessment system providing region-selective metric  $\Phi$  and predicted MOS  $MOS_{\Phi}$ .

**Metric 3:** In [3], a metric is reported that computes a structural similarity (SSIM) index between two images as

$$SSIM = \frac{(2\mu_R\mu_D + C_1)(2\sigma_{RD} + C_2)}{(\mu_R^2 + \mu_D^2 + C_1)(\sigma_R^2 + \sigma_D^2 + C_2)} \quad (6)$$

where  $\mu_R, \mu_D$  and  $\sigma_R, \sigma_D$  denote mean intensity and contrast of images  $I_R(x, y)$  and  $I_D(x, y)$ , respectively.  $C_1$  and  $C_2$  are constants used to avoid instabilities for very small  $\mu$  or  $\sigma$ .

**Metric 4:** Finally, the well known peak signal-to-noise ratio (PSNR) measures the fidelity difference of two image signals  $I_R(x, y)$  and  $I_D(x, y)$  on a pixel-by-pixel basis as

$$PSNR = 10 \log \frac{\eta^2}{MSE} \quad (7)$$

where  $\eta$  is the maximum pixel value, here 255. The mean square error is given as

$$MSE = \frac{1}{XY} \sum_{x=1}^X \sum_{y=1}^Y [I_R(x, y) - I_D(x, y)]^2 \quad (8)$$

where  $X$  and  $Y$  denote horizontal and vertical image dimensions, respectively.

### 3.2. Region-selective objective image quality metrics

In the following, the objective image quality metrics from Section 3.1 have been used to independently assess the image quality of ROI and BG to enable region-selective quality metric design. An overview of the region-selective quality prediction system is given in Fig. 2. The ROI is identified in the reference image  $I_R$  based on the corresponding mean ROI from the subjective experiment. Hence, prediction errors through automated ROI detection algorithms [1] are excluded and do not affect the region-selective quality metric design. ROI and BG extraction is then performed on the reference images  $I_R \in \mathcal{I}_R$  and distorted images  $I_D \in \{\mathcal{I}_B, \mathcal{I}_W\}$ . An ROI quality metric  $\Phi_{ROI}$  is calculated on the images  $I_{R,ROI}$  and  $I_{D,ROI}$ . Similar,  $I_{R,BG}$  and  $I_{D,BG}$  are used to assess the BG quality by computing  $\Phi_{BG}$ . In a pooling stage,  $\Phi_{ROI}$  and  $\Phi_{BG}$  are combined to a region-selective metric as

$$\Phi(\omega, \kappa, \nu) = [\omega \cdot \Phi_{ROI}^{\kappa} + (1 - \omega) \cdot \Phi_{BG}^{\kappa}]^{\frac{1}{\nu}} \quad (9)$$

where  $\Phi(\omega, \kappa, \nu) \in \{\Delta_{NHQM}, RRIQA, SSIM, PSNR\}$ ,  $\omega \in [0, 1]$ , and  $\kappa, \nu \in \mathbb{Z}^+$ . For  $\kappa = \nu$ , the expression in (9) is also

known as the weighted LP-norm. However, it will be shown later that in some cases better quality prediction performance can be achieved by allowing for the parameters  $\kappa$  and  $\nu$  to have different values. Finally, an exponential function is used to map  $\Phi(\omega, \kappa, \nu)$  to predicted MOS as follows

$$MOS_{\Phi(\omega, \kappa, \nu)} = a \cdot e^{b \cdot \Phi(\omega, \kappa, \nu)} \quad (10)$$

where  $a$  and  $b$  are derived from curve fitting of  $\Phi(\omega, \kappa, \nu)$  with  $\mathcal{M}_B$ . The exponential character of  $MOS_{\Phi(\omega, \kappa, \nu)}$  has been shown to account well for non-linearities in the HVS [4].

## 4. METRIC DESIGN AND EVALUATION

The region-selective metric design comprised of two parts; training (T) and validation (V). The training was performed using images  $\mathcal{I}_B$  and MOS  $\mathcal{M}_B$  from BIT subjective experiments. The pooling function parameters  $(\omega, \kappa, \nu)$  and exponential mapping parameters  $(a, b)$  obtained from the training are then used to compute the metrics on image set  $\mathcal{I}_W$  and validate their prediction performance using MOS  $\mathcal{M}_W$ . Training and validation was jointly conducted with respect to two aims: a) maximizing image quality prediction accuracy; b) maximizing generalization to unknown images. The former is evaluated using Pearson linear correlation coefficient  $\rho_P$  between MOS from subjective experiments and predicted MOS in (10). Further, Spearman rank order coefficient  $\rho_S$  is used to measure prediction monotonicity [6]. The generalization is evaluated using the absolute distance  $\Delta_{\rho_P} = |\rho_{P,T} - \rho_{P,V}|$  between the Pearson linear correlations on training and validation set. A smaller  $\Delta_{\rho_P}$  relates to a better generalization. All combinations of the pooling function parameters  $(\omega, \kappa, \nu)$  were taken into account for metric design. However, no noticeable improvements in prediction performance could be observed for values of  $\kappa$  and  $\nu$  larger than 5. Figure 3 shows the Pearson correlations for all metrics over weights  $\omega$ , for training and validation, and most favorable parameter set  $(\kappa, \nu, a, b)$ . One can see that the curves have very different characteristics. Therefore, the weights for the proposed metrics were individually assessed and selected as follows.

For  $\Delta_{NHQM}$  it occurs that  $\rho_{P,T}$  and  $\rho_{P,V}$  are very low where distance  $\Delta_{\rho_P}$  is smallest. Therefore, the weight was chosen for maximum  $\rho_{P,V}$  to maximize prediction accuracy, at the cost of reduced generalization. On the other hand, for  $RRIQA$  the weight was chosen with respect to maximum

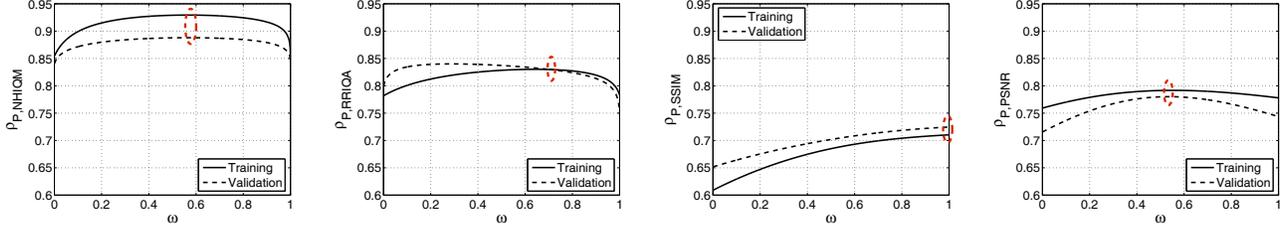


Fig. 3. Pearson linear correlations  $\rho_P$  for training and validation of the region-selective image quality metrics.

Table 2. Region-selective metric parameters.

	$\omega$	$\kappa$	$\nu$	$a$	$b$
$MOS_{NHIQM}$	0.567	3	5	116.145	-2.296
$MOS_{RRIQA}$	0.706	5	5	136.763	-0.205
$MOS_{SSIM}$	1	4	2	26.224	1.148
$MOS_{PSNR}$	0.522	1	5	0.204	2.855

generalization at only small cost of prediction accuracy. Finally, for both  $SSIM$  and  $PSNR$  the maxima of  $\rho_{P,T}$  and  $\rho_{P,V}$  and the minimum of  $\Delta_{\rho_P}$  coincide at the same weight allowing for maximum prediction accuracy and generalization. For illustration, the weights are marked with an ellipse in Fig. 3. Taking the above design issues into account, the parameters providing the best compromise between prediction accuracy and generalization are listed in Table 2. One can see that all metrics achieve the best performance for  $\omega > 0.5$ . Similar observations were made when using a linear pooling function with  $\kappa = \nu = 1$ . This confirms our earlier conjecture that structural degradations in the ROI have more severe impact on perceptual quality than degradations in the BG.

The prediction performance measures, corresponding to the parameters in Table 2, are shown in Table 3. In addition to the region-selective metrics  $\Phi$ , we computed whole image metrics  $\Theta$  and derived corresponding  $MOS_{\Theta}$  from curve fitting. These values are benchmarks to evaluate if the region-selective design is favorable for the considered metrics. One can see that prediction accuracy and monotonicity is enhanced for all metrics but in particular for  $SSIM$ , which is based on structural similarities. In addition, generalization to unknown images is improved for  $RRIQA$ ,  $SSIM$ , and  $PSNR$ .

## 5. CONCLUSIONS

In this paper, structural degradations in ROI and BG were independently assessed to enable region-selective image quality metric design. For this purpose, a subjective experiment was conducted to identify ROI for a set of natural images. Region-selective metrics were trained and validated on data from two independent subjective quality experiments. It has been shown that the prediction accuracy is significantly improved by using the region-selective quality metric design as compared to quality prediction on the whole image.

Table 3. Prediction accuracy and monotonicity.

		$\rho_{P,T}$	$\rho_{P,V}$	$\Delta_{\rho_P}$	$\rho_{S,T}$	$\rho_{S,V}$
$MOS_{NHIQM}$	$\Theta$	0.905	0.869	0.036	0.861	0.871
	$\Phi$	0.929	0.888	0.041	0.892	0.875
$MOS_{RRIQA}$	$\Theta$	0.769	0.848	0.079	0.677	0.823
	$\Phi$	0.83	0.83	0	0.752	0.797
$MOS_{SSIM}$	$\Theta$	0.6	0.638	0.038	0.461	0.612
	$\Phi$	0.71	0.725	0.015	0.582	0.683
$MOS_{PSNR}$	$\Theta$	0.778	0.741	0.037	0.644	0.632
	$\Phi$	0.792	0.78	0.012	0.695	0.751

## 6. REFERENCES

- [1] W. Osberger and A. M. Rohaly, "Automatic detection of regions of interest in complex video sequences," in *Proc. of SPIE HV & EI*, Jan. 2001, vol. 4299, pp. 361–372.
- [2] B. A. Wandell, *Foundations of Vision*, Sinauer Associates, Inc., 1995.
- [3] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. on Image Processing*, pp. 600–612, April 2004.
- [4] M. Kusuma, H.-J. Zepernick, and M. Caldera, "On the development of a reduced-reference perceptual image quality metric," in *Proc. of ICMCS*, Aug. 2005, pp. 178–184.
- [5] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," Rec. BT.500, 2002.
- [6] VQEG, "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, phase II," VQEG, Aug. 2003.
- [7] U. Engelke and H. J. Zepernick, "Quality evaluation in wireless imaging using feature-based objective metrics," in *Proc. of IEEE ISWPC*, Feb. 2007, pp. 367–372.
- [8] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," in *Proc. of SPIE HV & EI*, Mar. 2005, vol. 5666, pp. 149–159.