Electronic Research Archive of Blekinge Institute of Technology
<http://www.bth.se/fou/>

This is an author produced version of a conference paper. The paper has been peer-reviewed but may not include the final publisher proof-corrections or pagination of the proceedings.

Citation for the published Conference paper:

Title:

Author:

Conference Name:

Conference Year:

Conference Location:

Access to the published version may require subscription.

Published with permission from:

Optimal Region-Of-Interest Based Visual Quality Assessment

Ulrich Engelke and Hans-Jürgen Zepernick

Blekinge Institute of Technology, PO Box 520, 372 25 Ronneby, Sweden

ABSTRACT

Visual content typically exhibits regions that particularly attract the viewer's attention, usually referred to as regions-of-interest (ROI). In the context of visual quality one may expect that distortions occurring in the ROI are perceived more annoyingly than distortions in the background (BG). This is especially true given that the human visual system is highly space variant in sampling visual signals. However, this phenomenon of visual attention is only seldom taken into account in visual quality metric design. In this paper, we thus provide a framework for incorporation of visual attention into the design of an objective quality metric by means of region-based segmentation of the image. To support the metric design we conducted subjective experiments to both quantify the subjective quality of a set of distorted images and also to identify ROI in a set of reference images. Multiobjective optimization is then applied to find the optimal weighting of the ROI and BG quality metrics. It is shown that the ROI based metric design allows to increase quality prediction performance of the considered metric and also of two other contemporary quality metrics.

Keywords: Objective image quality metrics, region-of-interest, visual attention, subjective experiments, multi-objective optimization.

1. INTRODUCTION

Objectively measuring visually perceived quality has been subject to research for many years¹⁻³ and its significance has increasingly been recognised in recent years.⁴⁻⁹ Although by now there is a wide range of available objective quality metrics, most of them do not take into account that there are usually regions in visual content that particularly attract the viewers attention. This visual attention is an integral property of the human visual system (HVS) and the higher cognitive processing. To incorporate visual attention into image analysis may be of great interest with respect to visual quality assessment, since the viewer is usually more focused on regions-of-interest (ROI) and thus may perceive distortions within them more severe than in the background (BG). In addition, it is well known that the HVS is highly space variant in sampling and processing of visual signals, with the highest accuracy in the central point of focus, the fovea. Thus, one may assume that distortions in the BG are not perceived as severe as in the ROI. For these reasons it may be beneficial to incorporate visual attention into image quality metrics. The application range of metrics accounting for visual attention is broad and includes source codec optimization and unequal error protection (UEP) in wireless image or video communication, where the ROI may receive a stronger protection than the BG to improve the overall received quality.

A prerequisite of incorporating visual attention into quality metric design is to actually have means of detecting salient regions in the image. Various methodologies have been studied that allow to detect these regions in visual content. Very early work in this field has been conducted by Yarbus¹⁰ who did extensive subjective experiments using an eye tracker to analyse the gaze patterns of viewers. The algorithm by Privitera et al.¹¹ can predict spatial gaze patterns as were observed in eye tracking experiments, however, it was concluded that the sequential order of the pattern could not be predicted. Ninassi et al.¹² also utilised an eye tracker to create saliency maps and subsequently create simple distortion maps to quantify quality loss. Itti et al.¹³ have created a visual attention system with regards to the neuronal architecture of the early primate visual system where multiple scale image features are combined into a topographical saliency map. Maeder¹⁴ defines a formal approach for importance mapping and Osberger et al.¹⁵ utilise the outcomes of an eye tracker experiment to derive importance maps based on a set of factors that are known to influence visual attention. Similar

Further author information: (Send correspondence to U.E)

U.E: E-mail: ulrich.engelke@bth.se, Telephone: +46 (0)768 845 877

H.J.Z: E-mail: hans-jurg.zepernick@bth.se, Telephone: +46 (0)708 782 680

factors have been used by Pinneli et al.¹⁶ and are subject to a Bayesian learning approach to determine the likelihood of perceived interest for each of the factors. De Vleeschouwer et al.¹⁷ determine a level of interest for particular image regions using fuzzy modeling techniques. The above approaches have in common that they provide elaborate saliency information, for instance in terms of importance maps. Although this information would be highly valuable in many applications, such as image segmentation and content based image retrieval, in the context of UEP in wireless imaging one may rather have a less involved region distinction. In this context, Liu et al.¹⁸ deploy a probabilistic framework consisting of an appearance model and a motion model to discover and track ROI in video. Despite fairly high reliability of the algorithm, prediction errors may still be expected. Considering the above we decided to conduct a subjective ROI experiment instead, in which viewers were asked to select a ROI in a set of reference images. Thus, we can exclude prediction errors from automated algorithms which would bias the subsequent ROI based quality metric design.

We have earlier proposed an objective visual quality metric¹⁹ based on extraction of structural features. The aim of this previous work was to design a metric that can quantify the quality loss during image communication rather than the absolute quality of the received image. This may in turn be useful for radio resource management applications such as link adaptation, to assure a certain level of visual quality. Given the above, the metric was designed following the reduced-reference (RR) approach, meaning, that a set of low bandwidth features from the original image is extracted to assist the quality evaluation at the receiver. However, the metric performed the quality assessment uniformly over the whole image not taking the above mentioned visual attention into account. The aim of this paper now is to incorporate visual attention into the metric by means of a segmentation of the image into ROI and BG. It should be emphasized here, that the aim of the paper is not to design an automatic ROI detection algorithm, but rather to concentrate on the actual quality metric design. In practical applications, the metric may then be deployed in conjunction with automatic ROI detection algorithms. In this case, the metric is then computed independently in ROI and BG and subject to a weighted pooling into an overall quality metric. The weights for the ROI metric and BG metric account for their perceptual relevance and have been obtained using multiobjective optimization. We found that the incorporation of visual attention can improve the quality prediction accuracy and monotonicity of our previously proposed metric. In addition, we will show that the ROI based metric design framework could also successfully be deployed to other image quality metrics.

The paper is organised as follows. Our previously proposed objective quality metric is shortly summarised in Section 2 along with subjective quality experiments that we conducted to support the metric design. In Section 3 the subjective ROI experiment is explained, which we conducted to identify the ROI in the set of reference images. Section 4 then describes the ROI based metric design approach. The multiobjective optimization used to find the optimal parameters for ROI and BG is discussed in Section 5. Numerical results and an evaluation of the proposed ROI based metrics are provided in Section 6 and conclusions are finally drawn in Section 7.

2. NORMALIZED HYBRID IMAGE QUALITY METRIC

In this section we briefly summarise our previously proposed Normalised Hybrid Image Quality Metric (NHIQM).¹⁹ The metric was designed to be deployed for quality monitoring in image and video communication. Thus, the design followed the reduced-reference (RR) approach where low bandwidth information is extracted from the transmitted image and used at the receiver to measure quality loss during transmission. Subjective quality experiments that we conducted to support the metric design are reported in the following.

2.1 Subjective Quality Experiments

Visual quality can most precisely be judged by humans themselves and although subjective quality assessment is not applicable where real-time processing is required, the outcomes in terms of mean opinion scores (MOS) are an essential tool to support the design of objective image quality metrics. These metrics can in turn be applied for automated quality assessment. We thus conducted subjective image quality experiments in two independent laboratories to support the design of NHIQM. The first experiment took place at Blekinge Institute of Technology (BIT) in Ronneby, Sweden, and the other at the Western Australian Telecommunications Research Institute (WATRI) in Perth, Australia.²⁰ In each experiment 30 non-expert viewers were asked to judge the quality of distorted images as compared to the corresponding reference images. The experiment procedures were designed according to ITU-R Rec. BT.500-11.²¹

Table 1. Overview of the structural feature metrics, $\tilde{f}_i$, the corresponding artifacts that are measured, and the references to the reported algorithms.

Structural features		Corresponding artifacts	Reference
$\tilde{f}_1$	Block boundary differences Image activity	Blocking (Blur)	Wang et al. ²²
$\tilde{f}_2$	Edge smoothness	Blur	Marzilliano et al. ²³
$\tilde{f}_3$	Edge-based image activity	Ringing, (blocking)	Saha et al. ²⁴
$\tilde{f}_4$	Gradient-based image activity	Ringing, (blur)	Saha et al. ²⁴
$\tilde{f}_5$	Image histogram statistics	Intensity masking	Kusuma et al. ²⁵

The test images used in the experiments incorporated distortions caused by a simulation model of a wireless link, including source coding, channel coding, modulation, and a Rayleigh flat fading channel with additive white Gaussian noise. Therefore, a set $\mathcal{I}_R$ of 7 reference monochrome images of dimensions 512×512 pixels was chosen to account for different textures and complexities. The images were encoded into Joint Photographic Experts Group (JPEG) format. Two sets $\mathcal{I}_B$ and $\mathcal{I}_W$ of 40 distorted images each, were then created to be used in the BIT and WATRI experiments, respectively. The wireless link model resulted in a wide range of artifacts, in particular, blocking, blur, ringing, and intensity masking artifacts were observed in the test images in different degrees of severity.

The experiments at BIT and WATRI resulted in two respective sets of MOS, $\mathcal{M}_B$ and $\mathcal{M}_W$. For the metric design and validation we randomly created two sets of images, a training set $\mathcal{I}_T$ and a validation set $\mathcal{I}_V$. The training set contains 60 images, 30 from each $\mathcal{I}_B$ and $\mathcal{I}_W$, and the validation set contains the residual 20 images. Accordingly, we created corresponding MOS training set $\mathcal{M}_T$ and validation set $\mathcal{M}_V$.

2.2 Structural Feature Extraction

In order to capture the severity of the observed structural artifacts, and subsequently incorporate them into the design of NHIQM, we selected a set of objective structural feature metrics. Given the context of image communication, the feature metrics were selected with respect to three properties; the ability to quantify the corresponding structural artifact, the computational complexity, and a small numerical representation to keep the overhead low. An overview of the deployed feature metrics $\tilde{f}_i$ is given in Table 1, along with the corresponding artifacts and the references. The brackets indicate, that these artifacts are only measured indirectly by the corresponding feature metric.

In order to obtain a defined feature space, the features were scaled using an extreme value normalisation²⁶ as follows

$$f_{i,k} = \frac{\tilde{f}_{i,k} - \min_{k=1,\dots,K} \{\tilde{f}_{i,k}\}}{\delta_i}, \quad i = 1, \dots, I \quad (1)$$

where the denominator is given as

$$\delta_i = \max_{k=1,\dots,K} \{\tilde{f}_{i,k}\} - \min_{k=1,\dots,K} \{\tilde{f}_{i,k}\} \quad (2)$$

Here, K denotes the total number of training images, $\mathcal{I}_T$, validation images, $\mathcal{I}_V$, and reference images, $\mathcal{I}_R$. Resulting from the normalisation, we have $\forall i, k : 0 \leq f_{i,k} \leq 1$.

2.3 Feature Pooling

To further reduce the representation of the structural metric, the feature metrics are pooled in a single NHIQM value as follows

$$NHIQM = \sum_{i=1}^I w_i \cdot f_i \quad (3)$$

Here, the weights w_i were optimized regarding the impact of the corresponding artifact on the overall quality metric. Finally, to measure structural degradation between a distorted (d) image and its corresponding reference (r) image, we define an absolute difference

$$\Delta_{NHIQM} = |NHIQM_d - NHIQM_r| \quad (4)$$

This allows us to measure quality degradations induced during image communication rather than an absolute quality measure at the receiver. Given the single value for NHIQM, this is done using a minimal overhead in terms of RR information.

2.4 Exponential Mapping

In order to account for the non-linear quality processing in the human visual system one can further deploy a prediction function to map Δ_{NHIQM} to a predicted MOS as follows

$$MOS_{NHIQM} = a e^{b \Delta_{NHIQM}} \quad (5)$$

We have found that the exponential character of the prediction function accounts particularly well for the non-linear processing in the HVS.⁹

3. SUBJECTIVE ROI EXPERIMENT

The identification of salient regions in visual content is crucial in order to enable the incorporation of visual attention into the objective metric design. This task can be performed using the various methods as have been discussed in the Introduction. However, since many of these methods are not yet entirely reliable, an expected ROI prediction error may cause a bias in the objective quality metric design. For this reason we decided to conduct a subjective ROI experiment instead in which human observers have to vote for a ROI within the set of reference images $\mathcal{I}_R$ used in the quality experiments (see Section 2.1). The experiment procedures and evaluation are discussed in the following sections.

3.1 Experiment Procedures

We conducted the subjective ROI experiment at BIT and, as with the quality experiments, we had 30 non-expert viewers who participated. The viewers were presented one training image, two stabilisation images, and the seven reference images in $\mathcal{I}_R$. The task was then to select a region within each of the images that drew the viewers attention. In this context, the training image was used to explain the simple selection process, the stabilisation images allowed for the viewer to adapt to the process, and the actual test images in $\mathcal{I}_R$ were used for the evaluation. We did not put any restrictions on the size of the ROI to be selected, however, for simplicity we only considered rectangular regions.

3.2 Experiment Evaluation

The outcomes of the experiment enabled us to identify a subjective ROI for each image in $\mathcal{I}_R$ and ultimately to incorporate visual attention into Δ_{NHIQM} . In the following the experiment results are analysed in more detail.

3.2.1 Subjective ROI selections

The 30 ROI selections that we obtained for each reference image can be seen in Fig. 1. For better visualisation the ROI selections have been scaled by a factor 1/8 in both horizontal and vertical directions. As one would expect, faces strongly drew the attention of the viewers and were thus often selected as the ROI. However, the size of the area in the image that is covered by the face seems to play an important role. If a whole person is shown in the image (for instance Barbara), then the whole face is mostly chosen as the ROI. On the other hand, if most of the image is covered by the face (for instance Mandrill or Tiffany), then often details in the face are chosen rather than the whole face. In case of Mandrill, such details mainly comprised of the eyes and the nose, whereas for Tiffany, along with the eyes the mouth was chosen most frequently.

In case of a more complex scene, such as Peppers, the agreement on a ROI between the viewers is by far less pronounced as in the case where a human or a human face is shown. Here, different viewers have chosen different

Figure 1. All 30 ROI selections for the reference images in $\mathcal{I}_R$. For better visualisation the ROI have been scaled with a factor $1/8$.

peppers as ROI or selected the three big peppers in the center of the images. The disagreement is even worse in case of a natural scene such as Goldhill. Here, varying single houses have been selected frequently as well as the whole block of houses. Additionally, the little man walking down the street seemed to attract the attention of many viewers.

3.2.2 Outlier elimination

It was further observed that for all seven reference images there were some ROI selections that were far away from the majority of the votes. In other words, the x - and/or y -coordinate of the center of these ROI selections were numerically distant from the respective mean coordinates. These so called outliers were eliminated by adopting the criterium defined in²⁷ as follows

$$|x_C - \mu_{x_C}| > 2 \cdot \sigma_{x_C} \quad \text{or} \quad |y_C - \mu_{y_C}| > 2 \cdot \sigma_{y_C} \quad (6)$$

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, μ and σ 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

$$r_0 = \frac{R_0}{R} \quad (7)$$

where R_0 is the number of eliminated ROI selections and R the number of all ROI selections.

3.2.3 Mean ROI

Despite the variability of ROI selections in some of the images (see Section 3.2.1), we decided to only define one ROI for each of the reference images. The main reason for this is that many of the ROI selections are overlapping or even include each other. In case of the Tiffany image, for instance, people mostly chose the eyes, the mouth, or the whole face. Thus, selecting the face as ROI includes both eyes and mouth. Similar observations were made for the other images. Considering the above, we defined one ROI for each image as the mean of all 30 ROI selections. In particular, the location of the ROI was computed as the mean of all center coordinates in both x -direction and y -direction. The size of the ROI was computed as the mean dimensions in both x -direction and

Figure 2. Mean ROI for the images in $\mathcal{I}_R$ (black frame: before outlier elimination; brightened area: after outlier elimination).

y -direction. The mean ROI are shown in Fig. 2. Here, the black frame denotes the mean ROI before outlier elimination and the bright area indicates the mean ROI after outlier elimination (see Section 3.2.2). It should be noted, that in order for NHIQM to produce meaningful results, the ROI were adjusted to fall into the closest 8×8 block borders produced by the discrete cosine transform of the JPEG coder. However, the maximum error due to this necessary adjustment is only 0.78%.

3.2.4 Statistical analysis

A more detailed statistical analysis of the experiment outcomes is summarised in Table 2. Here, subscripts x_C and y_C denote the ROI center coordinates in x -direction and y -direction, respectively. Similarly, subscripts Δx and Δy denote the ROI dimensions in both x -direction and y -direction. Furthermore, μ and σ denote the mean and standard deviation over all 30 selections. All values are given in pixels with the origin in the lower left corner of the image. In addition, the outlier ratios r_0 are provided for all images.

The numerical results from the Table confirm the previous observations derived from Fig. 1 and Fig. 2. One can for instance see that the Barbara image exhibited the most outliers, which is why the mean ROI after outlier elimination is shifted more as compared to the other images. We believe that this is due to the location of the ROI which is, unlike for most of the other images, in the periphery of the image. In relation to that it is interesting to observe the high agreement (and thus low variance) for the x -direction of the center point, σ_{x_C} , in the Mandrill image. This can be attributed to the symmetry of the image and the fact that the major ROI selections (face, nose, eyes) were located on the symmetrical axis. Further observations from the Table in relation to Fig. 1 and Fig. 2 may be left to the reader.

4. ROI BASED METRIC DESIGN

In this section we will discuss the methodology we used to extend Δ_{NHIQM} , which assesses quality on the whole image, to a ROI based quality metric.

4.1 System Overview

An overview of the ROI based image quality assessment system is shown in Fig. 3. Here, both the reference image I_R and the distorted image I_D are segmented into ROI images, $I_{R,ROI}$ and $I_{D,ROI}$, and into BG images,

Table 2. Statistical analysis of the ROI experiment (x_C , y_C : ROI center point coordinates; Δx : ROI width; Δy : ROI height; μ : Mean over all 30 selections; σ : Standard deviation over all 30 selections; r_0 : Outlier ratio).

	μ_{x_C}	σ_{x_C}	μ_{y_C}	σ_{y_C}	$\mu_{\Delta x}$	$\sigma_{\Delta x}$	$\mu_{\Delta y}$	$\sigma_{\Delta y}$	r_0
Barbara	350	99.92	344	89.51	187.57	78.574	227.9	121.37	0.167
Elaine	260	42.24	263	46.96	225.53	105.31	202.6	132.45	0.1
Goldhill	288	65.36	204	87.79	215.83	137.69	204.6	107.86	0.1
Lena	278	60.2	227	31.99	202.1	90.035	235.53	141.92	0.1
Mandrill	256	8.27	339	86.14	293.9	58.444	277.07	159.15	0.033
Pepper	235	84.15	262	58.85	217.17	90.914	341.43	118.44	0.1
Tiffany	316	33.1	231	52.95	224.67	81.18	219.53	134.69	0.033

$I_{R,BG}$ and $I_{D,BG}$. The ROI identification is then performed on the reference image rather than the distorted image, since artifacts may disturb the ROI identification. In the context of image communication, the information about the ROI then needs to be transmitted along with the image to allow for the ROI and BG segmentation at the receiver. This is another reason why the shape of the ROI was kept simple since for a rectangular ROI the overhead in information to be transmitted is very low, as compared to more complex ROI or even multiple ROI. For the actual ROI identification we used the outcomes of the subjective experiment, however, automatic ROI detection algorithms may be deployed instead in a practical application. In case of NHIQM the structural features are then independently extracted in ROI ($f_{R,ROI}$, $f_{D,ROI}$) and BG ($f_{R,BG}$, $f_{D,BG}$) to compute the corresponding NHIQM values in both ROI ($NHIQM_{R,ROI}$, $NHIQM_{D,ROI}$) and BG ($NHIQM_{R,BG}$, $NHIQM_{D,BG}$). Eventually, structural degradations are measured independently in ROI and BG by computing $\Delta_{NHIQM,ROI}$ and $\Delta_{NHIQM,BG}$, respectively. To obtain an overall quality metric that incorporates visual attention (VA), $\Delta_{NHIQM,VA}$, the ROI metric and the BG metric are pooled in a final stage.

4.2 Pooling of ROI and Background

For the pooling of the ROI metric, $\Delta_{NHIQM,ROI}$, and the BG metric, $\Delta_{NHIQM,BG}$, we have chosen a variant of the well known Minkowski metric²⁸ as follows

$$\Delta_{NHIQM,VA}(\omega, \kappa, \nu) = [\omega \cdot \Delta_{NHIQM,ROI}^\kappa + (1 - \omega) \cdot \Delta_{NHIQM,BG}^\nu]^\frac{1}{\nu} \quad (8)$$

with $\omega \in [0, 1]$ and $\kappa, \nu \in \mathbb{R}^+$. For $\kappa = \nu$, the expression in (8) is also known as the weighted Minkowski metric. However, we have found that better quality prediction performance can be achieved by allowing for the parameters κ and ν to have different values. The weights ω regulate the impact of the $\Delta_{NHIQM,ROI}$ and $\Delta_{NHIQM,BG}$ on the overall quality metric $\Delta_{NHIQM,VA}$. With regards to our earlier conjecture that artifacts in the ROI may be perceived more annoying than in the background, one would expect the weight ω to have a value > 0.5 . The procedure to find the optimal weights will be discussed in the following section.

5. MULTIOBJECTIVE OPTIMIZATION

The optimal parameters ω_{opt} , κ_{opt} , and ν_{opt} can be obtained by means of optimization. In general, optimization is concerned with minimisation of an objective function, subject to a set of decision variables. Our objective was to maximise the correlation coefficient between $\Delta_{NHIQM,VA}$ and the MOS from the subjective experiment. However, we found that by doing so the metric worked very well on the training set of images but rather poorly on a test set of images. Thus we incorporated a second objective into the optimization which allows for better generalisation ability of the metric. We refer to this as a multiobjective optimization (MOO) problem, which is concerned with optimization of multiple, often conflicting objectives.²⁹ Two objectives are said to be conflicting when a decrease in one objective leads to an increase in the other. A MOO problem can be transformed into a single objective optimization, for instance by defining an objective as a weighted sum of multiple objectives. However, it is recommended to preserve the full dimensionality of the MOO problem.³⁰ The aim is then to find the optimal compromise between the objectives where system design aspects need to be taken into account to decide for the best trade-off solution.²⁹

Figure 3. System overview of the ROI based image quality assessment.

5.1 Definition of Multiple Objectives

Considering the above, we conduct a MOO based on a decision matrix $\mathbf{d} = [\omega \ \kappa \ \nu] \in \mathbb{D} \subset \mathbb{R}^3$. The MOO is conducted with respect to two objectives: a) maximising image quality prediction accuracy O_A and b) maximising generalisation performance O_G . Objective O_A defines the metrics ability to predict MOS with minimal error and is measured as the Pearson linear correlation between metric $\Delta_{NHIQM,VA}$ and MOS $\mathcal{M}$ on the training set

$$\rho_P = \frac{\sum_k (\Delta_{NHIQM,VA,k} - \bar{\Delta}_{NHIQM,VA})(\mathcal{M}_k - \bar{\mathcal{M}})}{\sqrt{\sum_k (\Delta_{NHIQM,VA,k} - \bar{\Delta}_{NHIQM,VA})^2} \sqrt{\sum_k (\mathcal{M}_k - \bar{\mathcal{M}})^2}} \quad (9)$$

where $\bar{\Delta}_{NHIQM,VA}$ and $\bar{\mathcal{M}}$, respectively, denote the mean values of $\Delta_{NHIQM,VA}$ and $\mathcal{M}$. As mentioned before, optimizing the weights using only objective O_A would likely overtrain the metric, meaning, it would work very well on the training set but not on a set of unknown images. Therefore, objective O_G defines the metrics ability to perform quality prediction on a set of unknown images. We compute it as the absolute difference of ρ_P on training and validation set as follows

$$\Delta\rho_P = |\rho_{P,T} - \rho_{P,V}| \quad (10)$$

We thus define the objective vector as

$$\mathbf{O}(w) = \begin{pmatrix} O_A(w) \\ O_G(w) \end{pmatrix} = \begin{pmatrix} -|\rho_{P,T}| \\ \Delta\rho_P \end{pmatrix} \quad (11)$$

The decision matrix $\mathbf{d}$ is evaluated by assigning it an objective vector $\mathbf{O}$ in the objective space $\mathbb{O}$: $\mathbb{D} \rightarrow \mathbb{O} \subset \mathbb{R}^2$.

5.2 Goal Attainment Method

We determine the optimal solution using the goal attainment method.³¹ Here, goals $\mathbf{O}^* = (O_A^* \ O_G^*)^T$ are specified, which can be interpreted as the desired level of the corresponding objectives $\mathbf{O} = (O_A \ O_G)^T$. This requires sufficient intuitive understanding of the problem to know what values one would like to attain for each of the objectives. We define the MOO problem as

$$P_G : \begin{cases} \min & z \\ \text{s.t.} & \mathbf{O}(w) - \boldsymbol{\lambda} \cdot z \leq \mathbf{O}^* \end{cases} \quad (12)$$

where the magnitude of $\boldsymbol{\lambda} = (\lambda_A \ \lambda_G)^T$ determines how close the objectives $(O_A(w) \ O_G(w))^T$ are to the goals $(O_A^* \ O_G^*)^T$. It is typically set to the absolute value of the goals

$$\lambda_A = |O_A^*| \quad \text{and} \quad \lambda_G = |O_G^*| \quad (13)$$

The quantity $\boldsymbol{\lambda} \cdot z$ corresponds to the degree of under- or overattainment of the goals $\mathbf{O}^*$. With regards to results from our earlier metric design,¹⁹ we define the goals as $O_A^* = -0.9$ and $O_G^* = 0.0001$. If we relax the generalisation goal O_G^* we can significantly increase the prediction accuracy $\rho_{P,T}$ on the training set. However, this would be done at the cost of prediction accuracy $\rho_{P,V}$ on the validation set. Since we want to avoid such overfitting of the metric to the training data, the generalisation goal O_G^* is chosen to be very small.

Table 3. Optimal ROI based metric parameters.

	ω_{opt}	κ_{opt}	ν_{opt}
$\Delta_{NHIQM,VA}$	0.593	3.142	4.066
$SSIM_{VA}$	0.823	4.062	0.534
VIF_{VA}	0.978	2.928	0.798

6. NUMERICAL RESULTS

In order to investigate if the ROI based metric design is also applicable to other quality metrics we chose two contemporary image quality metrics for which we also conducted the ROI based metric design and optimization, as outlined in Sections 4 and 5. In particular, the Structural Similarity (SSIM) index³² and the Visual Information Fidelity (VIF) criterion³³ were extended to incorporate visual attention and are denoted as $SSIM_{VA}$ and VIF_{VA} . In this section we evaluate the optimization outcomes and the quality prediction performance of all three considered metrics.

6.1 Optimal Parameters

The optimal parameters for $\Delta_{NHIQM,VA}$, $SSIM_{VA}$, and VIF_{VA} are shown in Table 3. It is interesting to note that the weights ω are larger than 0.5 for all three metrics, which confirms our earlier conjecture that the ROI metrics should receive a higher weight due to artifacts in the ROI being more annoying than in the background. Also, one can see that the optimal parameters for $SSIM_{VA}$ and VIF_{VA} are fairly similar, meaning, that both have a ω_{opt} at the higher end of the scale and a significantly larger value for κ_{opt} as compared to ν_{opt} . This is somewhat not unexpected since it has been shown³⁴ that both metrics have very strong relationships in their methodologies of objectively assessing perceived quality.

6.2 Quality Prediction Performance of the ROI Based Metrics

The quality prediction performance of the ROI based metrics has been evaluated using the Pearson linear correlation coefficient and the Spearman rank order correlation as measures of prediction accuracy and monotonicity, respectively.²⁷ The results are shown in Fig. 4-6.

In Fig. 4 it can be seen that the prediction accuracy of Δ_{NHIQM} can be improved by about 5% by incorporating visual attention. This accounts for both the training and the validation set of images. It can also be observed that the prediction accuracy on both training and validation set is very similar which is due to the generalisation objective that prevented the metric from being overtrained. In addition to the improvement of prediction accuracy one can also see an improvement in prediction monotonicity.

The quality prediction performance for SSIM is shown in Fig. 5. It can be seen that a strong improvement can be achieved for both the prediction accuracy and monotonicity and on both the training and validation set. Similar observations can be made for VIF, shown in Fig. 6. Considering the above, we have shown that the framework for ROI based metric design, which we propose in this paper, does not only improve the quality prediction performance of our metric Δ_{NHIQM} but can also be applied to other image quality metrics. Given the simplicity of the design, the additional computational complexity and overhead in terms of RR information are minimal as compared to the actual gain in quality prediction performance.

6.3 An Illustrative Example

To further illustrate the quality prediction performance improvement of the considered metrics, two distorted images are shown in Fig. 7 along with the corresponding reference image. Here, the image in the middle contains artifacts mainly located in the ROI whereas the image to the right contains artifacts in the BG (upper right corner). The related objective quality metrics and the subjective MOS are given in Table 4. In addition to the three metrics we have also provided the predicted MOS MOS_{NHIQM} which includes the non-linear quality processing, as discussed in Section 2.4. All metrics are given for both the computation on the whole image and the ROI based metric design. It should be noted that for Δ_{NHIQM} a higher value relates to lower quality whereas for the other three metrics, a higher value indicates higher quality.

Figure 4. Pearson linear correlation coefficient ρ_P and Spearman rank order correlation ρ_S for Δ_{NHIQM} and $\Delta_{NHIQM,VA}$ on both training and validation set.

Figure 5. Pearson linear correlation coefficient ρ_P and Spearman rank order correlation ρ_S for SSIM and SSIM_{VA} on both training and validation set.

Figure 6. Pearson linear correlation coefficient ρ_P and Spearman rank order correlation ρ_S for VIF and VIF_{VA} on both training and validation set.

Given that the artifacts in both distorted images are of similar severeness, the whole image metrics do not distinguish between differently perceived qualities. This is not in line with the MOS for which the observers have voted the quality of the ROI distorted image (middle) about 10% lower as compared to the BG distorted image (right). On the other hand, the ROI based metrics reflect much better the quality difference between the two distorted images by incorporating the spatial locality of the artifacts into the metric design.

Table 4. Comparison of subjective MOS with the ROI based quality metrics and the related metrics as computed on the whole image.

Artifact location	Whole image metrics				ROI based metrics				MOS
	Δ_{NHIQM}	MOS_{NHIQM}	SSIM	VIF	$\Delta_{NHIQM,VA}$	$MOS_{NHIQM,VA}$	$SSIM_{VA}$	VIF_{VA}	
ROI	0.095	72.713	0.97	0.969	0.158	59.101	0.656	0.758	49.6
BG	0.094	72.884	0.972	0.969	0.098	72.081	0.716	0.928	59.833

Figure 7. Illustrative example regarding the annoyance of artifacts in the ROI as compared to the background (left: reference image; middle: ROI distorted image; right: BG distorted image).

7. CONCLUSIONS

In this paper, we proposed a framework to incorporate visual attention into image quality metric design by means of spatial segmentation of the image into ROI and BG. A subjective ROI experiment has been conducted and evaluated in order to allow for ROI identification in a set of reference images. Multiobjective optimization has been applied to find optimal parameters for the ROI based pooling function with respect to maximal prediction accuracy while at the same time good generalisation ability of the metric. It has been shown that the ROI based metric design can significantly increase the prediction accuracy and monotonicity of three image quality metrics, while imposing only very little additional computational complexity and overhead in terms of RR information.

REFERENCES

- [1] Mannos, J. and Sakrison, D., “The effects of a visual fidelity criterion on the encoding of images,” *IEEE Trans. on Information Theory* **20**(4), 525–536 (1974).
- [2] Lukas, F. and Budrikis, Z., “Picture quality prediction based on a visual model,” *IEEE Trans. on Communications* **30**(7), 1679–1692 (1982).
- [3] Daly, S., “Visible differences predictor: an algorithm for the assessment of image fidelity,” *Human Vision, Visual Processing, and Digital Display III* **1666**, 2–15 (1992).
- [4] Watson, A. B., Hu, J., and McGowan, J. F., “Digital video quality metric based on human vision,” *SPIE Journal of Electronic Imaging* **10**(1), 20–29 (2001).
- [5] Martens, J. B., “Multidimensional modeling of image quality,” *Proc. of the IEEE* **90**(1), 133–153 (2002).
- [6] Keelan, B. W., [*Handbook of Image Quality: Characterization and Prediction*], CRC Press (2002).
- [7] Pappas, T. N., Safranek, R. J., and Chen, J., “Perceptual criteria for image quality evaluation,” in [*Handbook of Image and Video Processing*], Bovik, A. C., ed., ch. 8.2, 939–960, Academic Press (2005).
- [8] H. R. Wu and K. R. Rao (Ed.), [*Digital Video Image Quality and Perceptual Coding*], CRC Press (2006).
- [9] Engelke, U., “A perceptual-based objective quality metric for wireless imaging.” Lic. dissertation, Blekinge Institute of Technology (2008).
- [10] Yarbus, A. L., [*Eye Movements and Vision*], Plenum (1967).

- [11] Privitera, C. M. and Stark, L. W., "Algorithms for defining visual regions-of-interest: Comparison with eye fixations," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **22**(9), 970–982 (2000).
- [12] Ninassi, A., Meur, O. L., Callet, P. L., and Barba, D., "Does where you gaze on an image affect your perception of quality? Applying visual attention to image quality metric," in [*Int. Conf. on Image Processing*], *Proc. IEEE* **2**, 169–172 (2007).
- [13] Itti, L., Koch, C., and Niebur, E., "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **20**(11), 1254–1259 (1998).
- [14] Maeder, A. J., "The image importance approach to human vision based image quality characterization," *Pattern Recognition Letters* **26**(3), 347–354 (2005).
- [15] Osberger, W. and Rohaly, A. M., "Automatic detection of regions of interest in complex video sequences," in [*Human Vision and Electronic Imaging*], *Proc. IS&T/SPIE* **4299**, 361–372 (2001).
- [16] Pinneli, S. and Chandler, D. M., "A Bayesian approach to predicting the perceived interest of objects," in [*Int. Conf. on Image Processing*], *Proc. IEEE*, 2584–2587 (2008).
- [17] Vleschouwer, C. D., Marichal, X., Delmot, T., and Macq, B., "A fuzzy logic system able to detect interesting areas of a video sequence," in [*Human Vision and Electronic Imaging II*], *Proc. IS&T/SPIE* **3016**, 234–245 (1997).
- [18] Liu, D. and Chen, T., "DISCOV: A framework for discovering objects in video," *IEEE Trans. on Multimedia* **10**(2), 200–208 (2008).
- [19] Engelke, U. and Zepernick, H. J., "Pareto optimal weighting of structural impairments for wireless imaging quality assessment," in [*Int. Conf. on Image Processing*], *Proc. IEEE*, 373–376 (2008).
- [20] Kusuma, T. M., Zepernick, H.-J., and Caldera, M., "On the development of a reduced-reference perceptual image quality metric," in [*Int. Conf. on Multimedia Communications Systems*], 178–184 (2005).
- [21] International Telecommunication Union, "Methodology for the subjective assessment of the quality of television pictures," Rec. BT.500-11, ITU-R (2002).
- [22] Wang, Z., Sheikh, H. R., and Bovik, A. C., "No-reference perceptual quality assessment of JPEG compressed images," in [*Int. Conf. on Image Processing*], *Proc. IEEE* **1**, 477–480 (2002).
- [23] Marziliano, P., Dufaux, F., Winkler, S., and Ebrahimi, T., "A no-reference perceptual blur metric," in [*Int. Conf. on Image Processing*], *Proc. IEEE* **3**, 57–60 (2002).
- [24] Saha, S. and Vemuri, R., "An analysis on the effect of image features on lossy coding performance," *IEEE Signal Processing Letters* **7**(5), 104–107 (2000).
- [25] Kusuma, T. M., *A Perceptual-based Objective Quality Metric for Wireless Imaging*, PhD thesis, Curtin University of Technology, Perth, Australia (2005).
- [26] Ohm, J.-R., [*Multimedia Communication Technology: Representation, Transmission and Identification of Multimedia Signals*], Springer (2004).
- [27] Video Quality Experts Group, "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, phase II." VQEG (2003).
- [28] de Ridder, H., "Minkowski-metrics as a combination rule for digital-image-coding impairments," in [*Human Vision, Visual Processing, and Digital Display III*], *Proc. IS&T/SPIE* **1666**, 16–26 (1992).
- [29] Fonseca, C. M. and Fleming, P. J., "Multiobjective optimization," in [*Evolutionary Computation 2: Advanced Algorithms and Operations*], Back, T., Fogel, D. B., and Michalewicz, Z., eds., ch. 5, 25–37, Taylor & Francis (2000).
- [30] Zadeh, L. A., "Optimality and non-scalar-valued performance criteria," *IEEE Trans. on Automatic Control* **8**(1), 59–60 (1963).
- [31] Gembicki, F. W. and Haimes, Y. Y., "Approach to performance and sensitivity multiobjective optimization: The goal attainment method," *IEEE Trans. on Automatic Control* **20**(6), 769–771 (1975).
- [32] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. on Image Processing* **13**(4), 600–612 (2004).
- [33] Sheikh, H. R. and Bovik, A. C., "Image information and visual quality," *IEEE Trans. on Image Processing* **15**(2), 430–444 (2006).
- [34] Seshadrinathan, K. and Bovik, A. C., "Unifying analysis of full reference image quality assessment," in [*Int. Conf. on Image Processing*], *Proc. IEEE*, 1200–1203 (2008).