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

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

Citation for the published Journal paper:

Title:

Author:

Journal:

Year:

Vol.

Issue:

Pagination:

URL/DOI to the paper:

Access to the published version may require subscription.

Published with permission from:

Framework for optimal region of interest–based quality assessment in wireless imaging

Ulrich Engelke

Hans-Jürgen Zepernick

Blekinge Institute of Technology

P.O. Box 520

372 25 Ronneby

Sweden

ulrich.engelke@bth.se

Abstract. Images usually exhibit regions that particularly attract the viewer's attention. These regions are typically referred to as regions of interest (ROI), and the underlying phenomenon in the human visual system is known as visual attention (VA). In the context of image quality, one can expect that distortions occurring in the ROI are perceived as being more annoying compared to distortions in the background. However, VA is seldom taken into account in existing image quality metrics. In this work, we provide a VA framework to extend existing image quality metrics with a simple VA model. The performance of the framework is evaluated on three contemporary image quality metrics. We further consider the context of wireless imaging where a broad range of artifacts can be observed. To facilitate the VA-based metric design, we conduct subjective experiments to both obtain a ground truth for the subjective quality of a set of test images and to identify ROI in the corresponding reference images. A methodology is further discussed to optimize the VA metrics with respect to quality prediction accuracy and generalization ability. It is shown that the quality prediction performance of the three considered metrics can be significantly improved by deploying the proposed framework. © 2010 SPIE and IS&T. [DOI: 10.1117/1.3267097]

1 Introduction

Mean opinion scores (MOS) obtained in subjective image quality experiments are to date the only widely accepted measures of perceived visual quality.¹ On the other hand, image fidelity metrics such as the peak signal-to-noise ratio (PSNR) are still predominantly used as objective metrics, even though they are well known to correlate poorly with human perception of quality. For this reason, the efforts to find objective metrics that can predict subjectively rated quality have been increased in recent years,^{2–7} where many methods are based on or related to early efforts in modeling the human perception of visual quality.^{8–10} Although 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 viewer's attention. This phenomenon, referred to as visual attention (VA),¹¹ is an integral property of the human visual system (HVS) and higher cognitive processing deployed to

reduce the complexity of scene analysis.¹² For this purpose, a subset of the available visual information is selected by scanning the visual scene and focusing on the most salient regions.¹³ Incorporating a VA model into image quality assessment is thus of great importance, since the viewer may be more likely to detect artifacts in the salient regions, typically referred to as regions of interest (ROI), as compared to regions of low saliency, here referred to as 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, and strongly diminishing accuracy toward the periphery of the visual field. As such, artifacts in the ROI may be perceived in more detail and consequently as being more annoying than in the BG.

This is particularly true in applications where artifacts are found to be not just uniformly distributed over the whole image but also clustered in certain areas of the scene. For instance, source coding artifacts are usually more uniformly distributed than artifacts that can be observed in a wireless communication system where the hostile nature of the wireless channel causes a broad range of artifact types and severities. However, most of the existing metrics consider only source coding artifacts and artificial noise as distortions. In this work, we focus on the context of a wireless imaging scenario, including the integral parts of a wireless link such as source coding, channel coding, modulation, and the wireless channel. We propose a framework to incorporate a simple VA model into existing image quality metrics. The framework is nonintrusive, meaning that it can be readily applied to existing image quality metrics without changing the actual metric. The application range of quality metrics accounting for this VA framework is broad, including 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.

In the following sections we discuss in more detail VA modeling, in particular the detection of salient regions in visual scenes, and we summarize the proposed framework.

Paper 09064SSPR received Apr. 30, 2009; revised manuscript received Jun. 27, 2009; accepted for publication Jul. 22, 2009; published online Jan. 7, 2010. This paper is a revision of a paper presented at the SPIE conference on Human Vision and Electronic Imaging, January 2009, San Jose, California. The paper presented there appears (unrefereed) in SPIE Proceedings Vol. 7240.

Fig. 1 Overview of the proposed framework.

1.1 Visual Attention Modeling and Salient Region Identification

In the context of quality metric design, VA models¹⁴ play a vital role in identifying salient regions in the visual scene. Many models follow early works such as the feature integration theory by Treisman and Gelade,¹⁵ the guided search by Wolfe, Cave, and Franzel,¹⁶ or the neural-based architecture by Koch and Ullman.¹⁷ In general, two processes affect VA, known as bottom-up attention and top-down attention. The former is a rapid, saliency-driven, and task independent process, whereas the latter is slower, volition-controlled, and task dependent.¹³ Typically, VA models aim to predict either bottom-up or top-down VA by either following a HVS-related approach or a content-based approach. HVS related methods are based on modeling various properties of the HVS such as multiscale processing, contrast sensitivity, and center surround processing. On the other hand, content-based methods model different visual factors that are known to attract attention such as object color, shape, and location.

Various models have been proposed in the literature aiming toward the detection of salient regions in an image. Very frequently, these models are developed and validated based on visual fixation patterns, as they can be obtained through eye tracking experiments. Early work in this field has been conducted by Yarbus,¹⁸ who did extensive subjective experiments using an eye tracker to analyze the gaze patterns of a number of viewers. Privitera and Stark¹⁹ proposed an algorithm that was able to predict spatial gaze patterns as obtained in eye tracking experiments. It was concluded, however, that the sequential order of the pattern could not be predicted. Ninassi *et al.*²⁰ also utilized an eye tracker to create saliency maps and subsequently create simple distortion maps to quantify quality loss. Itti, Koch, and Niebur¹³ created a VA 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. Another HVS-based VA system has been proposed by Le Meur *et al.*,²¹ which builds saliency maps based on a three-stage model including a visibility, a perception, and a grouping stage. Maeder²² defines a formal approach for importance mapping, and Osberger and Rohaly²³ utilize the outcomes of an eye tracker experiment to derive importance maps based on a number of factors that are known to influence VA. Similar factors have been used by Pinnelli and Chandler,²⁴ and are subject to a

Bayesian learning approach to determine the likelihood of perceived interest for each of the factors. De Vleeschouwer *et al.*²⁵ determined a level of interest for particular image regions using fuzzy modeling techniques.

What the prior approaches have in common is that they provide elaborate saliency information, for instance, in terms of visual fixation patterns and importance maps. Although this information would be highly valuable in many applications, such as image segmentation and content-based image retrieval, there are other applications for which one may rather have a less involved description of the saliency information. For instance, for UEP in wireless imaging, a simple saliency description would be preferable to facilitate the assignment of different channel codes for the purpose of varying protection levels according to the perceptual relevance of a region. A simple saliency description would further keep the computational complexity and overhead, in terms of side information, at a decent level. In this context, Liu and Chen²⁶ deployed a simple 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.

1.2 Proposed Framework

The framework proposed in this work is based on the work that we presented in Ref. 27. The basic idea is to include a simple VA model into existing image quality metrics that do not consider any saliency information, and as a result, improve the metrics' quality prediction performance. An overview of the framework is shown in Fig. 1. The first step is the identification of a ROI in the reference image I_R . The ROI coordinates are then used to segment both the undistorted reference image I_R and a distorted version of it I_D into ROI images $I_{R,ROI}$ and $I_{D,ROI}$, and BG images $I_{R,BG}$ and $I_{D,BG}$. An image quality metric Φ , is then independently computed on the ROI and BG images, resulting in a quality metric for the ROI, Φ_{ROI} , and one for the BG, Φ_{BG} . In this work we consider three different quality metrics. Finally, a pooling function is deployed to determine a single quality metric Φ_{VA} , incorporating the simple VA model based on ROI and BG segmentation. The parameters of the pooling function are optimized independently for each of the considered metrics.

In a practical application, one may deploy automated algorithms and models, as discussed in the previous sec-

Fig. 2 Simulation model of a wireless link.

tion, to facilitate online ROI detection. However, to avoid ROI detection errors and subsequent errors in the metric design, we conducted a subjective experiment instead in which human observers identified the ROI in a set of reference images. It should be noted here that, like gaze patterns from eye tracking experiments, such an ROI selection process is one way of obtaining a ground truth for salient regions in a visual scene. In recent work,²⁸ we found that the locations of the selected ROI strongly correlate with visual fixation patterns (VFP) that we obtained in eye tracking experiments on the same set of reference images. This applies especially for the first couple of fixations after appearance of the image, which may indicate that the ROI selections reflect better the saliency-driven, bottom-up attention.

In this work, we show that the incorporation of VA using the previous outlined framework allows for improving the quality prediction accuracy and monotonicity of the considered metrics. It should be emphasized here again that the framework does not require the code of an existing metric to be changed, since the metrics are independently computed in their original form on both ROI and BG. It is necessary to identify the ROI; however, it should be emphasized here that the aim of the work is not to design an automatic ROI detection algorithm, but rather to concentrate on the actual quality metric design. For this reason, we conducted the subjective experiment for ROI identification. In the context of image communication, the information about the ROI location and size needs to be transmitted along with the image to allow for the ROI and BG segmentation at the receiver. To keep the transmission overhead (in terms of side information about the ROI) low, it is desirable to remain a simple complexity of the ROI.

The work is organized as follows. In Sec. 2 we discuss our previous work on wireless imaging quality assessment and introduce briefly two subjective image quality experiments that we conducted to support the metric design. In Sec. 3 we describe and analyze in detail a subjective ROI experiment, which we conducted to identify the ROI in a set of reference images. The three image quality metrics considered here for the VA framework are then shortly introduced in Sec. 4. The pooling of ROI and BG metrics is discussed in Sec. 5, along with the optimization method deployed to find the optimal pooling parameters. Numerical results and an evaluation of the proposed ROI-based metrics are provided in Sec. 6 and conclusions are finally drawn in Sec. 7.

2 Wireless Imaging Quality

The integral parts of a wireless link model are shown in Fig. 2. At the transmitter, source encoding, channel encoding, and modulation are applied to the image, and at the receiver the inverse operations are deployed. In the following, the wireless link model is outlined, as we used it to create a number of test images. These test images were subsequently presented in two subjective image quality experiments that we conducted.

2.1 Wireless Link Model

In the scope of this work, we consider a particular setup of the wireless link model as outlined before. To be precise, the Joint Photographic Experts Group (JPEG) format has been chosen to source encode the images. JPEG is a lossy image coding technique using a block discrete cosine transform (DCT)²⁹-based algorithm. Due to the quantization of DCT coefficients, artifacts such as blocking and blur may be introduced during source encoding. A (31,21) Bose-Chaudhuri-Hocquenghem (BCH)³⁰ code was then used to encode all 21 information bits into 31 code bits to enhance the error resilience of the image prior to transmission over the error prone channel. Finally, binary phase shift keying (BPSK) was deployed for modulation. An uncorrelated Rayleigh flat fading channel in the presence of additive white Gaussian noise (AWGN) was implemented as a simple model of the wireless channel.³¹ To produce severe transmission conditions, the average bit energy to noise power spectral density ratio E_b/N_0 was chosen as 5 dB. These conditions may cause bit errors or burst errors in the transmitted signal, which are beyond the correction capabilities of the channel decoder, and as a result, artifacts may be induced in the decoded image in addition to the ones purely caused by the lossy source encoding.

2.2 Test Images

A set $\mathcal{I}_R$ of seven well-known monochrome reference images, namely Barbara (B), Elaine (E), Goldhill (G), Lena (L), Mandrill (M), Peppers (P), and Tiffany (T) of dimensions 512×512 pixels, was chosen to account for different textures and complexities. The wireless link model outlined in Sec. 2.1 was then deployed to create two sets $\mathcal{I}_1$ and $\mathcal{I}_2$ of 40 test images each to be presented in the two subjective quality experiments. The specific setup of the model resulted in test images that covered a broad range of artifact types and severities. In particular, blocking, blur, ringing, intensity masking, and noise artifacts were observed in the

Fig. 3 Examples of test images as presented in the quality experiments: (a) Barbara with blocking and ringing; (b) Elaine with ringing; (c) Lena with blocking, intensity masking, and noise; (d) Tiffany with in-block blur and local blocking; (e) Mandrill with severe blocking; (f) Peppers with ringing, intensity masking (brighter), and blocking; and (g) Goldhill with intensity masking (darker) and ringing.

test images in different degrees of severity and in various combinations. Some examples of test images are shown in Fig. 3 to illustrate the range of artifacts induced into the images by the wireless link model.

2.3 Subjective Image Quality Experiments

MOS obtained in subjective image quality experiments are widely accepted as a ground truth for the design and validation 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, which are explained in detail in Ref. 7 and is briefly summarized in the following.

The first experiment (E1) took place at the Blekinge Institute of Technology (BIT) in Ronneby, Sweden. 30 non-expert viewers participated, of which 24 were male and 6 were female. The second experiment (E2) was conducted at the Western Australian Telecommunications Research Institute (WATRI) in Perth, Australia.³² Again, 30 nonexpert viewers participated, of which 25 were male and 5 were female. The procedures of both experiments were designed according to ITU-R Rec. BT.500-11.³³ In both experiments, two viewers participated in parallel in each session. The images in E1 were presented on two Sony CPD-E200 17-in. cathode ray tube (CRT) monitors, and in E2 on a pair of 17-in. CRT monitors of type Dell and Samtron 75E. The viewing distance was chosen as four times the height of the test images. The double stimulus continuous quality scale (DSCQS) was used as the assessment method in which the test images are presented in an alternating order with the corresponding reference images. Each alternation lasted 3 sec with a 2-sec midgray screen in between. During the last two alternations, the viewers were asked to rate the quality of both images on a continuous scale from 0 to 100 with 100 being the best quality. Five labels (Excellent, Good, Fair, Poor, and Bad) along the continuous scale were further provided to assist the viewers with the quality rating. Prior to the actual test images, the viewers were presented four training images for us to explain the assessment process, and also five stabilization images for the viewers to adapt to the process. The test images in $\mathcal{I}_1$ were then presented in experiment E1, whereas the test images in $\mathcal{I}_2$ were presented in experiment E2. To counteract viewers' fatigue, each session was split into two sections with a break in between.

The experiments at BIT and WATRI resulted in two sets of MOS, $\mathcal{M}_1$ and $\mathcal{M}_2$, corresponding to the image sets $\mathcal{I}_1$ and $\mathcal{I}_2$, respectively. The MOS covered the whole range of subjective qualities from Bad to Excellent, in accordance to the broad range of artifact severities, and represent the basis on which the objective metrics can be designed and validated. 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 $\mathcal{I}_1$ and $\mathcal{I}_2$, and the validation set contains the residual 20 images. Accordingly, we created the corresponding MOS training set $\mathcal{M}_T$ and validation set $\mathcal{M}_V$.

3 Subjective Region of Interest

The identification of salient regions in visual content is crucial to enable the incorporation of visual attention into the objective metric design. However, a ground truth regarding the location and extent of the salient regions is needed, similar to the MOS from subjective quality experiments. This task can be performed using the various methods discussed in Sec. 1.1. 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 had to select a ROI within the set of reference images, $\mathcal{I}_R$, used in the quality experiments. The experiment procedures and evaluation are discussed in the following sections.

3.1 Experiment Procedures

We conducted the subjective ROI experiment at BIT. As with the quality experiments, we had 30 nonexpert viewers who participated, of which 17 were male and 13 were female. The viewers were presented a number of images on a 19-in. DELL display at a viewing distance of four times the height of the test images. The viewer's task was to select a region within each of the images that drew most of their attention. We presented one training image to explain the simple selection process and two stabilization images for the viewer to adapt to the selection process. The viewers were then presented the seven reference images in $\mathcal{I}_R$. We did not put any restrictions on the size of the ROI to be selected other than that the selected region needed to be a subset of the whole image. For simplicity, we considered only rectangular-shaped ROI and allowed for only one ROI selection per image. We further allowed the viewers to re-select a ROI in case of dissatisfaction with the selected ROI. We did not impose any limits regarding the time needed for the ROI selection; however, given the simplicity of the ROI selection process, most viewers were able to conduct the experiment within a few minutes.

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 deploy the ROI-based metric design framework as proposed in this work. In the following, the experiment results are analyzed in detail.

3.2.1 Subjective region of interest selections

The 30 ROI selections that we obtained for each reference image are visualized in Fig. 4. Here, all ROI selections have been added to the image as an intensity shift and as such, a brighter area indicates more overlapping ROI and thus a higher saliency in that particular region. To enhance the visualization of the ROI, the images have been darkened before adding the ROI.

As one would expect, faces strongly drew the attention of the viewers and were thus primarily 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 the 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 the 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 present. Here, different viewers have chosen different peppers as ROI or selected the three big peppers in the

Fig. 4 All 30 ROI selections for each of the reference images in $\mathcal{I}_R$. The images have been darkened for better visualization of the overlaid ROI.

center of the image. Most attention has actually been drawn by the two stems of the peppers, which may be due to their prominent appearance on the otherwise fairly uniform skins of the peppers. The disagreement between viewers is even larger in the case of a natural scene, such as Goldhill. Here, varying single houses have been selected frequently as well

Fig. 5 Mean μ over all 30 ROI selections for: (a) ROI center coordinates, and (b) ROI horizontal (x coordinate) and vertical (y coordinate) dimensions.

as the whole block of houses. Additionally, the little man walking down the street seemed to be of interest to many viewers.

3.2.2 Statistical analysis

To gain more insight into the characteristics of the ROI selections, we further analyze the ROI locations and ROI dimensions using simple statistics, such as the mean μ and the standard deviation σ . The results for the mean μ are summarized in Fig. 5, and for the standard deviation σ in Fig. 6. Here, x denotes the horizontal coordinate and y the vertical coordinate with the origin being in the bottom left corner of the image. Furthermore, x_C and y_C denote the ROI center coordinates and x_Δ and y_Δ denote the ROI dimensions in the x and y directions, respectively. The labels on the abscissa denote the first letters of the reference images in $\mathcal{I}_R$ (see Sec. 2.2).

In Fig. 5(a) it can be seen that the mean of the ROI center coordinates, x_C and y_C , are around the image center for most of the images. This may be somewhat expected, since the salient region is typically placed toward the center of a natural scene when, for instance, taking a photograph. The only exception here is the Barbara image, for which the mean ROI is significantly shifted to the upper right corner toward the face. It is also worth noting that x_C for the image Mandrill lies exactly in the horizontal center of the image, which can be explained by the axis of symmetry of the Mandrill face being centrally located in the horizontal direction.

Figure 5(b) reveals that the mean ROI dimensions for most images are very similar in both x and y directions. Interestingly, the Mandrill image reveals much larger di-

Fig. 6 Standard deviation σ over all 30 ROI selections for: (a) ROI center coordinates, and (b) ROI horizontal (x coordinate) and vertical (y coordinate) dimensions.

mensions, which is caused by many viewers selecting the whole face or the nose as ROI of considerable size. The large extent of the y coordinate in the case of the Peppers image is due to many selections of either all three big peppers or selections of the long pepper on the left.

The standard deviation of the ROI center coordinates in Fig. 6(a) reveals information about the agreement of the viewers as to where the ROI is located, similar to confidence intervals with regard to MOS in subjective quality experiments. In this respect, a larger standard deviation and thus a lower agreement indicates that there may be either no dominant ROI or that there are multiple ROI present in the visual content. Given the previous, the small values in the cases of Elaine, Lena, and Tiffany further support earlier observations (see Sec. 3.2.1) that faces are of strong interest to the viewers and that the agreement between viewers is high. On the other hand, larger standard deviations such as for Goldhill and Peppers indicate that the identification of a dominant ROI is not as clear, and thus that the agreement between the viewers is lower. An exception is again given by the Barbara image, which comprises a face but has, on the contrary, also the highest standard deviations. This may be due to the face being located in the periphery of the image and also due to other objects being present that some viewers found of interest, such as the object on the table to the left. With respect to the Mandrill image, it is interesting to point out the difference between the standard deviations in the x and y directions. One can see that there is strong agreement that the ROI is located on the horizontal center of the image; however, the agreement is low as to the vertical location of the ROI. This was also observed in the visual inspection of the ROI where many

Table 1 Outlier ratios for the reference images in $\mathcal{I}_R$.

Image	Barbara	Elaine	Goldhill	Lena	Mandrill	Peppers	Tiffany
r_0	5/30	3/30	3/30	3/30	1/30	3/30	1/30

selections were found for the eyes, nose, and the whole face, all of them being located on the horizontal center but spread in the vertical direction.

Finally, comparing Figs. 6(b) and 6(a) reveals that the disagreement between viewers regarding the size of the ROI seems to be large compared to the disagreement about location. It is further observed that for all images, apart from Goldhill, the disagreement is considerably higher in the vertical direction (y coordinate) as compared to the horizontal direction (x coordinate). This may be due to the viewers selecting either a whole body, a face, or parts of a face, where in all cases the width of the ROI selection is not as much affected as the height. This accounts in particular for images such as Barbara, Lena, Mandrill, and Tiffany.

3.2.3 Outlier elimination

In addition to the prior observations, we found 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 coordinates of the center of these ROI selections were numerically distant from the respective mean coordinates. We eliminated these so-called outliers by adopting the criterion defined by the Video Quality Experts Group in³⁴ as follows,

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

As such, a ROI is considered to be an outlier if the distance of either x_C and/or y_C to the respective mean over all 30 selections is at least twice the corresponding standard deviation. Based on the number of eliminated outliers, we define an outlier ratio for each of the images as

$$r_0 = \frac{R_0}{R}, \quad (2)$$

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

The outlier ratios for all images are summarized in Table 1. One can see that the Barbara image exhibited the most outliers, which we believe is due to the location of the ROI in the periphery of the image. The least outliers can be observed for the Mandrill and Tiffany image, which are also the images with the face being present to a larger extent as compared to the other face images. Hence, no other objects are present in the visual scene that may distract the viewers' attention away from the face.

3.2.4 Mean region of interest

Despite the variability of ROI selections in some of the images (see Sec. 3.2.1), we decided to only define one ROI for each of the reference images. The reasons for this decision are three-fold. First, and most importantly, many of the

ROI selections overlap or even include each other. For instance, in the case of the Tiffany image, 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. Second, in the context of wireless imaging, we aim to keep the overhead and computational complexity low. Since a higher number of deployed ROI is directly related to an increased overhead in terms of side information and also an increased complexity in terms of the number of computed metrics, we decided on only one ROI. Last, deploying only a single ROI is in agreement with the subjective experiment in which we asked the viewers to select a single ROI.

Considering this, we defined one ROI for each image as the mean over all 30 ROI selections. In particular, the location of the ROI was computed as the mean over all center coordinates x_C and y_C . The size of the ROI was computed as the mean over x_Δ and y_Δ . The mean ROI are shown in Fig. 7. Here, the black frame denotes the mean ROI before outlier elimination, and the bright area indicates the mean ROI after outlier elimination (see Sec. 3.2.3).

3.2.5 Segmentation into region of interest image I_{ROI} and background image I_{BG}

The mean ROI coordinates after outlier elimination were used to segment all reference and distorted images into ROI images I_{ROI} and BG images I_{BG} . In particular, the ROI images were obtained by cutting out the area according to the mean ROI center coordinates μ_C , and the mean ROI dimensions μ_Δ (see Fig. 5). The BG images then comprised of the remainder of the images with the pixels in the ROI set to zero.

4 Objective Image Quality Metrics

In the following sections we briefly introduce the three image quality metrics that we consider in this work. All three metrics were designed to assess the quality uniformly over the whole image, not taking into account VA to salient regions in the visual scene. Within the framework proposed in this work, each of the metrics is applied on both the ROI, I_{ROI} , and the BG, I_{BG} , independently (see Fig. 1). As such, no modifications of the actual metrics need to be performed, allowing seamless deployment of the framework to existing image quality metrics.

4.1 Normalized Hybrid Image Quality Metric

We previously proposed the normalized hybrid image quality metric (NHIQM),³⁵ which was designed to evaluate quality degradations in a wireless imaging system. Here, a set of objective structural feature metrics was deployed to measure blocking, blur, ringing, and intensity masking artifacts. Given the context of image communication, the feature metrics were selected with respect to three properties;

Fig. 7 Mean ROI for the reference images in $\mathcal{I}_R$ (black frame: before outlier elimination; brightened area: after outlier elimination).

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 feature metrics f_i and the corresponding artifacts is given in Table 2.^{36–39} The feature metrics are then pooled in a single NHIQM value given by

Table 2 Overview of the feature metrics f_i , the corresponding artifacts, and the references to the reported algorithms.

Structural features		Corresponding artifacts	Reference
f_1	Block boundary differences	Blocking	Ref. 36
f_2	Edge smoothness	Blur	Ref. 37
f_3	Edge-based image activity	Ringing, noise	Ref. 38
f_4	Gradient-based image activity	Ringing, noise	Ref. 38
f_5	Image histogram statistics	Intensity masking	Ref. 39

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

which further reduces the numerical representation of the metric, and thus the overhead needed to transmit the reference information. The weights w_i in Eq. (3) regulate the impact of the corresponding artifact on the overall quality metric. The weights were optimized with respect to the metric's quality prediction accuracy and generalization ability, in a similar fashion as is outlined in Sec. 5.2. To measure structural degradation between a distorted (d) image and its corresponding reference (r) image, an absolute difference was further defined as

$$\Delta_{\text{NHIQM}} = |\text{NHIQM}_d - \text{NHIQM}_r|. \quad (4)$$

This allowed us to measure quality degradations induced during image communication rather than only absolute quality at the receiver. Finally, the nonlinear quality processing in the HVS is accounted for by further deploying a prediction function to map Δ_{NHIQM} to a predicted MOS as follows

$$\text{MOS}_{\text{NHIQM}} = a \exp(b \Delta_{\text{NHIQM}}), \quad (5)$$

where the parameters a and b are determined using curve fitting of Δ_{NHIQM} with the training set of MOS $\mathcal{M}_T$.⁷

4.2 Structural Similarity Index

The structural similarity (SSIM) index⁴⁰ is based on the assumption that the HVS is highly adapted to the extraction of structural information from a visual scene. As such, it predicts structural degradations between two images based on simple intensity and contrast measures. The final SSIM index is given by

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (6)$$

where μ_x, μ_y and σ_x, σ_y denote the mean intensity and contrast of image signals x and y , respectively. The constants C_1 and C_2 are used to avoid instabilities in the structural similarity comparison that can occur for certain mean intensity and contrast combinations ($\mu_x^2 + \mu_y^2 = 0, \sigma_x^2 + \sigma_y^2 = 0$).

4.3 Visual Information Fidelity Criterion

The visual information fidelity (VIF) criterion proposed in Ref. 41 approaches the image quality assessment problem from an information theoretical point of view. In particular, the degradation of visual quality due to a distortion process is measured by quantifying the information available in a reference image and the amount of this reference information that can be still extracted from the test image. As such, the VIF criterion measures the loss of information between two images. For this purpose, natural scene statistics, and in particular Gaussian scale mixtures (GSM) in the wavelet domain, are used to model the images. The proposed VIF metric is given by

$$\text{VIF} = \frac{\sum_{j \in \text{subbands}} I(\tilde{C}^{N,j}; \tilde{F}^{N,j} | \tilde{S}^{N,j})}{\sum_{j \in \text{subbands}} I(\tilde{C}^{N,j}; \tilde{E}^{N,j} | \tilde{S}^{N,j})}, \quad (7)$$

where $\tilde{C}$ denotes the GSM, N denotes the number of GSM used, and $\tilde{E}$ and $\tilde{F}$ denote the visual output of a HVS model, respectively, for the reference and test image.

5 Optimal Pooling of Region of Interest and Background Metrics

The metrics introduced in the previous section are used to independently assess the quality of the ROI and the BG in an image, as illustrated in Fig. 1. In the following section, a pooling function is discussed that was deployed to combine the ROI and BG metrics into a single quality metric that accounts for VA. An optimization methodology is further described that was implemented to find the optimal parameters of the pooling function.

5.1 Pooling of Region of Interest and Background Metrics

Let Φ be a general definition of an objective image quality metric, as already shown in Fig. 1. Given the metrics that we deploy within the scope of this work, we can then specify $\Phi \in \{\Delta_{\text{NHQM}}, \text{SSIM}, \text{VIF}\}$. Furthermore, let Φ_{ROI} be a metric computed on the ROI image I_{ROI} , and Φ_{BG} be a metric computed on the BG image I_{BG} . We then deploy a variant of the well-known Minkowski metric⁴² to obtain the final metric Φ_{VA} as follows

$$\Phi_{\text{VA}}(\omega, \kappa, \nu) = [\omega \cdot \Phi_{\text{ROI}}^\kappa + (1 - \omega) \cdot \Phi_{\text{BG}}^\nu]^{1/\nu}, \quad (8)$$

with $\Phi_{\text{VA}}(\omega, \kappa, \nu) \in \{\Delta_{\text{NHQM,VA}}, \text{SSIM}_{\text{VA}}, \text{VIF}_{\text{VA}}\}$, $\omega \in [0, 1]$, and $\kappa, \nu \in \mathbb{Z}^+$. For $\kappa = \nu$, the expression in Eq. (8) is also known as the weighted Minkowski metric. However, we have found that better quality prediction performance can be achieved by allowing the parameters κ and ν to have different values. The weights ω regulate the impact of the Φ_{ROI} and Φ_{BG} on the overall quality metric Φ_{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 parameters for ω , κ , and ν are discussed in the following section.

5.2 Multiobjective Optimization of ω , κ , and ν

The optimal parameters ω_{opt} , κ_{opt} , and ν_{opt} were obtained by means of optimization. In general, optimization is concerned with minimization of an objective function, subject to a set of decision variables. Our objective was to maximize the correlation coefficient between Φ_{VA} and the MOS $\mathcal{M}_T$ from the subjective experiment. However, we found that by doing so the metric worked very well on the training set of images $\mathcal{I}_T$ but rather poorly on the validation set of images $\mathcal{I}_V$. Thus we incorporated a second objective into the optimization that allows for better generalization ability of the metric. We refer to this as a multiobjective optimization (MOO) problem, which is concerned with optimization of multiple, often conflicting, objectives.⁴³ Two objectives are said to be conflicting when a decrease in one objective leads to an increase in the other. A MOO problem could 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.⁴⁴ The aim is then to find the optimal compromise between the objectives, where system design aspects need to be taken into account to decide the best trade-off solution.⁴³

5.2.1 Definition of multiple objectives

Considering the prior, we perform a MOO based on a decision vector $\mathbf{d} = [\omega \ \kappa \ \nu] \in \mathbb{D} \subset \mathbb{R}^3$. The MOO is conducted with respect to two objectives: 1. maximizing image quality prediction accuracy O_A , and 2. maximizing generalization performance O_G . Objective O_A defines the metric's ability to predict MOS with minimal error, and is measured as the Pearson linear correlation between metric Φ_{VA} and MOS $\mathcal{M}$ on the training set

$$\rho_P = \frac{\sum_k (\Phi_{\text{VA},k} - \bar{\Phi}_{\text{VA}})(\mathcal{M}_k - \bar{\mathcal{M}})}{\left[\sum_k (\Phi_{\text{VA},k} - \bar{\Phi}_{\text{VA}})^2 \right]^{1/2} \left[\sum_k (\mathcal{M}_k - \bar{\mathcal{M}})^2 \right]^{1/2}}, \quad (9)$$

where $\bar{\Phi}_{\text{VA}}$ and $\bar{\mathcal{M}}$, respectively, denote the mean values of Φ_{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 metric's ability to perform quality prediction on a set of unknown images. We compute it as the absolute difference of ρ_P on the training and validation set as follows

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

We thus define the objective vector as

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

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

5.2.2 Goal attainment method

We determine the optimal solution using the goal attainment method.⁴⁵ Here, goals $\mathbf{O}^* = (O_A^* \ O_G^*)^T$ are specified,

Table 3 Optimal parameters of the VA-based metrics Φ_{VA} .

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

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,³⁹ we define the goals as $O_A^*=-0.9$ and $O_G^*=0.0001$. If we relax the generalization 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 generalization goal O_G^* is chosen to be very small.

5.2.3 Optimal parameters

The optimal parameters for $\Delta_{\text{NHIQM},VA}$, 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 the 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 shown⁴⁶ that both metrics have very strong relationships in their methodologies of objectively assessing perceived quality.

6 Numerical Results

In the following sections we evaluate the quality prediction performance of all three considered metrics.

6.1 Quality Prediction Performance of the Visual Attention Based Metrics Φ_{VA}

The quality prediction performance of the ROI-based metrics has been evaluated using the Pearson linear correlation

Fig. 8 Pearson linear correlation coefficient ρ_P on both training and validation set for: (a) Δ_{NHIQM} and $\Delta_{\text{NHIQM},VA}$; (b) SSIM and SSIM_{VA} ; and (c) VIF and VIF_{VA} .

coefficient ρ_P and the Spearman rank order correlation ρ_S as measures of prediction accuracy and monotonicity, respectively.³⁴ The results for the Pearson correlation are presented in Fig. 8 and for the Spearman correlation in Fig. 9. The pale blue bars represent the correlations for the whole image metrics, Δ_{NHIQM} , SSIM , and VIF , and the dark blue bars for the metrics incorporating the VA model, $\Delta_{\text{NHIQM},VA}$, SSIM_{VA} , and VIF_{VA} . The results are further shown for both the training set and the validation set of images.

In Fig. 8 it can be seen that the prediction accuracy of the VA metrics Φ_{VA} could be improved as compared to the

Fig. 9 Spearman rank order correlation ρ_S on both training and validation set for: (a) Δ_{NHIQM} and $\Delta_{\text{NHIQM},VA}$; (b) SSIM and SSIM_{VA} ; and (c) VIF and VIF_{VA} .

whole image metrics Φ for all three considered metrics and on both the training and validation set. In particular, an improvement of about 5% can be observed for $\Delta_{\text{NHIQM},\text{VA}}$ on both the training and validation set. It can also be observed that ρ_P has similar values on both the training and validation set, thus maintaining the excellent generalization ability of Δ_{NHIQM} . The improvement in prediction accuracy is even more significant for SSIM_{VA} , for which ρ_P could be increased by about 15 to 20%. Here, the generalization of SSIM_{VA} is not quite as good, which may be due to SSIM already showing a lower correlation on the validation set as compared to the training set. Finally, VIF_{VA} shows both a significant improvement in prediction accuracy of about 12% as well as a well-maintained generalization.

Similar observations for ρ_P can also be done for ρ_S shown in Fig. 9. The Spearman rank order correlation is improved for all three metrics, $\Delta_{\text{NHIQM},\text{VA}}$, SSIM_{VA} , and VIF_{VA} , on both the training and validation set. However, the improvement seems to be less prevalent for the Pearson correlation, and also, the generalization ability is worse in the case of SSIM_{VA} and VIF_{VA} . Both phenomena can be explained due to the optimization being performed with respect to the Pearson correlation rather than the Spearman correlation.

Considering the previous, we have shown that the framework for ROI-based metric design was successfully deployed to three contemporary image quality metrics. Δ_{NHIQM} , SSIM, and VIF. From Figs. 8 and 9 one can infer that a higher gain in prediction performance can be expected for metrics that initially have a worse performance. To be more precise, Δ_{NHIQM} has the best performance on both the training and validation set, followed by VIF and SSIM. However, the improvement in both prediction accuracy and monotonicity have been highest for SSIM_{VA} , followed by VIF_{VA} and $\Delta_{\text{NHIQM},\text{VA}}$.

6.2 Illustrative Examples

To further illustrate the quality prediction performance improvement of the considered metrics, we take a closer look at the Lena and Tiffany images in Fig. 10. Here, the images in Figs. 10(a) and 10(c) contain mainly artifacts within the ROI determined from the ROI experiment (see Sec. 3). In the Lena image, one can observe several distorted rows crossing her face around the eyes and the nose, and in the Tiffany image we find distorted rows along the mouth and nose and some additional artifacts in the eyes. On the other hand, the images in Figs. 10(b) and 10(d), respectively, exhibit artifacts mainly located in the top right corner and in the bottom. As such, these artifacts are located in the BG of the image.

The related objective image quality metrics and the MOS for the four images in Fig. 10 are presented in Table 4. Here, we compare the metrics Φ , as computed in the whole images, to the VA based metrics Φ_{VA} . In the case of Δ_{NHIQM} , we have additionally provided the predicted MOS, $\text{MOS}_{\text{NHIQM}}$, which includes the nonlinear quality processing, as discussed in Sec. 4.1. 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.

Given that the artifacts in both distorted Lena images are of similar severeness, the whole image metrics Φ do not

Fig. 10 Illustrative examples regarding the annoyance of artifacts in the ROI compared to the background: (a) Lena with ROI artifacts; (b) Lena with BG artifacts; (c) Tiffany with ROI artifacts; and (d) Tiffany with BG artifacts.

distinguish between differently perceived qualities. In fact, the values of all four metrics are about the same for both the ROI distorted Lena image and the BG distorted Lena image. This is not in line with the MOS for which the observers have voted the quality of the ROI distorted image [Fig. 10(a)] about 10% lower as compared to the BG distorted image [Fig. 10(b)], indicating that the artifact in the ROI was perceived as more annoying. 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.

Similar observations can be made with regards to the Tiffany image. Here again, the ROI distorted image [Fig. 10(c)] has received a lower MOS as compared to the BG distorted image [Fig. 10(d)]. In this case, all four objective metrics Φ actually rated the ROI distorted image of better quality than the BG distorted image, which may be due to the BG artifacts being slightly more prevalent than the ROI artifacts. On the other hand, the VA-based metrics Φ_{VA} inversed the order of the quality rating, thus being more in line with the MOS.

7 Conclusions

We introduce a framework to extend existing image quality metrics with a simple VA model, based on a spatial image segmentation into ROI and BG. The model is nonintrusive, meaning that the image quality metrics do not need to be modified in any way. Instead, the metrics are computed independently on extracted ROI and BG images. A pooling function is then deployed to compute the final quality metric that accounts for VA. A methodology is further pre-

Table 4 Comparison of the VA based quality metrics, Φ_{VA} and the related metrics as computed on the whole image, Φ .

Artifact location		Whole image metrics Φ				VA based metrics Φ_{VA}				
		Δ_{NHIQM}	MOS_{NHIQM}	SSIM	VIF	$\Delta_{NHIQM,VA}$	$MOS_{NHIQM,VA}$	$SSIM_{VA}$	VIF_{VA}	MOS
Lena	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
Tiffany	ROI	0.084	72.101	0.971	0.956	0.183	54.331	0.683	0.62	47.27
	BG	0.116	66.576	0.97	0.939	0.096	72.537	0.912	0.972	55.33

sented to determine the optimal parameters of the pooling function with respect to quality prediction accuracy and generalization ability of the metric.

To obtain saliency information for our set of reference images, we conduct a subjective experiment in which human observers selected ROI in each of the images. As such, we based the subsequent metric design on a reliable ground truth. The ROI experiment is explained and analyzed in detail, providing valuable information about human preferences in visual scenes.

Three contemporary image quality metrics are considered in this work, which are all designed to assess quality uniformly over the whole image. It is shown that the VA framework significantly increases both the prediction accuracy and monotonicity of the metrics, while maintaining the generalization ability. Some example images further illustrate the capability of the developed VA metrics to account for the locality of the artifacts within the image. Given the simplicity of the methodology, only very little additional computational complexity and overhead in terms of ROI side information are imposed. This is particularly important in the scope of wireless communication, where computational power and channel bandwidth are scarce.

7.1 Limitations

The proposed framework may to some degree be dependent of the distortion model (see Sec. 2.1) and the considered set of reference images (see Sec. 2.2). An extension of the framework to different types of distortions and visual content would be straightforward by following the methodology as outlined in this work. However, such an extension requires obtaining both the MOS and the ROI coordinates from the respective subjective experiments and also a new set of test images. In this respect, it would be of further interest to consider images of different resolution to account for various screen sizes (for instance, mobile phones and laptops) and viewing distances. As such, one could determine if the image resolution has an impact on the outcome of the framework.

7.2 Future Work

The framework presented in this work is based on subjective ROI selections conducted exclusively on the reference images in $\mathcal{I}_R$. Hence, the degree to which the artifacts distract the viewers attention from the visual content is not taken into account in the ROI experiment. To gain a better understanding of the viewing behavior of human observers

when viewing distorted images, we thus conduct an eye tracking experiment at the University of Western Sydney, Australia, where we record the gaze patterns of the viewers when observing the test images in $\mathcal{I}_1$ and $\mathcal{I}_2$ and the reference images in $\mathcal{I}_R$. Based on the resulting gaze patterns, we investigate the degree to which the artifacts distract the attention from the visual content. Preliminary results promise to reveal interesting insights into the connections between the gaze patterns and the structural distortions. We discuss these results in Ref. 28, where we also analyze the relationship between the visual fixations from the eye tracking experiment and the ROI from the experiment explained in Sec. 3. We find that there is indeed a strong correlation between the locations of ROI and visual fixations. We will continue to evaluate the eye tracking experiment to gain more insight into the impact of wireless imaging artifacts on VA.

References

1. S. Winkler, *Digital Video Quality—Vision Models and Metrics*, John Wiley and Sons, New York (2005).
2. A. B. Watson, J. Hu, and J. F. McGowan, "Digital video quality metric based on human vision," *J. Electron. Imaging* **10**(1), 20–29 (2001).
3. J. B. Martens, "Multidimensional modeling of image quality," *Proc. IEEE* **90**(1), 133–153 (2002).
4. B. W. Keelan, *Handbook of Image Quality: Characterization and Prediction*, CRC Press, Boca Raton, FL (2002).
5. T. N. Pappas, R. J. Safranek, and J. Chen, "Perceptual criteria for image quality evaluation," in *Handbook of Image and Video Processing*, A. C. Bovik, Ed., Chaps. 8.2, 939–960, Academic Press, San Diego (2005).
6. *Digital Video Image Quality and Perceptual Coding*, H. R. Wu and K. R. Rao, Eds., CRC Press, Boca Raton, FL (2006).
7. U. Engelke, "A perceptual-based objective quality metric for wireless imaging," Licentiate Thesis, Blekinge Institute of Technology, Ronneby, Sweden (2008).
8. J. Mannos and D. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Trans. Inf. Theory* **20**(4), 525–536 (1974).
9. F. Lukas and Z. Budrikis, "Picture quality prediction based on a visual model," *IEEE Trans. Commun.* **30**(7), 1679–1692 (1982).
10. S. Daly, "Visible differences predictor: an algorithm for the assessment of image fidelity," *Proc. SPIE* **1666**, 2–15 (1992).
11. J. Wolfe, "Visual attention," in *Seeing*, K. K. D. Valois, Ed., pp. 335–386, Academic Press, San Diego (2000).
12. J. M. Wolfe and T. S. Horowitz, "What attributes guide the deployment of visual attention and how do they do it?," *Nat. Rev. Neurosci.* **5**, 1–7 (June 2004).
13. L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(11), 1254–1259 (1998).
14. L. Itti and C. Koch, "Computational modelling of visual attention," *Nat. Rev. Neurosci.* **2**, 192–203 (2001).
15. A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cogn. Psychol.* **12**, 97–136 (Jan. 1980).

16. J. M. Wolfe, K. R. Cave, and S. L. Franzel, "Guided search: an alternative to the feature integration model for visual search," *J. Exp. Psychol. Hum. Percept. Perform.* **15**, 419–433 (Aug. 1989).
17. C. Koch and S. Ullman, "Shifts in selection in visual attention: towards the underlying neural circuitry," *Hum. Neurobiol.* **4**(4), 219–227 (1985).
18. A. L. Yarbus, *Eye Movements and Vision*, Plenum, New York (1967).
19. C. M. Privitera and L. W. Stark, "Algorithms for defining visual regions-of-interest: comparison with eye fixations," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(9), 970–982 (2000).
20. A. Ninassi, O. L. Meur, P. Le Callet, and D. Barba, "Does where you gaze on an image affect your perception of quality? Applying visual attention to image quality metric," *Proc. IEEE* **2**, 169–172 (2007).
21. O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau, "A coherent computational approach to model bottom-up visual attention," *IEEE Trans. Pattern Anal. Mach. Intell.* **28**, 802–817 (May 2006).
22. A. J. Maeder, "The image importance approach to human vision based image quality characterization," *Pattern Recogn. Lett.* **26**(3), 347–354 (2005).
23. W. Osberger and A. M. Rohaly, "Automatic detection of regions of interest in complex video sequences," *Proc. SPIE* **4299**, 361–372 (2001).
24. S. Pinneli and D. M. Chandler, "A Bayesian approach to predicting the perceived interest of objects," *15th Int. Conf. on Image Processing (ICIP'08)*, pp. 2584–2587, IEEE, Piscataway, NJ (2008).
25. C. De Vleeschouwer, X. Marichal, T. Delmot, and B. Macq, "A fuzzy logic system able to detect interesting areas of a video sequence," *Proc. SPIE* **3016**, 234–245 (1997).
26. D. Liu and T. Chen, "DISCOV: a framework for discovering objects in video," *IEEE Trans. Multimedia* **10**(2), 200–208 (2008).
27. U. Engelke and H. J. Zepernick, "Optimal region-of-interest based visual quality assessment," *Proc. SPIE* **7240**, 724011 (Jan. 2009).
28. U. Engelke, H. J. Zepernick, and A. J. Maeder, "Visual attention modeling: region-of-interest versus fixation patterns," *Proc. 27th Picture Coding Symp.*, IEEE, Piscataway, NJ (2009).
29. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed., Prentice Hall, Englewood Cliffs, NJ (2002).
30. S. Lin and D. J. Costello, *Error Control Coding: Fundamentals and Applications*, Prentice Hall, Englewood Cliffs, NJ (1983).
31. A. F. Molisch, *Wireless Communications*, Wiley-IEEE Press, New York (2005).
32. T. M. Kusuma, H. J. Zepernick, and M. Caldera, "On the development of a reduced-reference perceptual image quality metric," in *Int. Conf. Multimedia Commun. Syst.*, pp. 178–184, IEEE, Piscataway, NJ (2005).
33. International Telecommunication Union, "Methodology for the subjective assessment of the quality of television pictures," Rec. BT.500-11, ITU-R, ITU, Geneva, Switzerland (2002).
34. 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, Geneva, Switzerland (2003).
35. U. Engelke and H. J. Zepernick, "Pareto optimal weighting of structural impairments for wireless imaging quality assessment," in *15th Int. Conf. on Image Processing (ICIP'08)*, pp. 373–376, IEEE, Piscataway, NJ (2008).
36. Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Int. Conf. on Image Processing (ICIP'02)*, Vol. 1, pp. 477–480, IEEE, Piscataway, NJ (2002).
37. P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *Int. Conf. on Image Processing (ICIP'02)*, Vol. 3, pp. 57–60, IEEE, Piscataway, NJ (2002).
38. S. Saha and R. Vemuri, "An analysis on the effect of image features on lossy coding performance," *IEEE Signal Process. Lett.* **7**(5), 104–107 (2000).
39. T. M. Kusuma, "A perceptual-based objective quality metric for wireless imaging," PhD Thesis, Curtin Univ. of Technology, Perth, Australia (2005).
40. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
41. H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.* **15**(2), 430–444 (2006).
42. H. De Ridder, "Minkowski-metrics as a combination rule for digital-image-coding impairments," *Proc. SPIE* **1666**, 16–26 (1992).
43. C. M. Fonseca and P. J. Fleming, "Multiobjective optimization," in *Evolutionary Computation 2: Advanced Algorithms and Operations*, T. Back, D. B. Fogel, and Z. Michalewicz, Eds., Chaps 5, 25–37, Taylor & Francis, New York (2000).
44. L. A. Zadeh, "Optimality and non-scalar-valued performance criteria," *IEEE Trans. Autom. Control* **8**(1), 59–60 (1963).
45. F. W. Gembicki and Y. Y. Haimes, "Approach to performance and sensitivity multiobjective optimization: the goal attainment method," *IEEE Trans. Autom. Control* **20**(6), 769–771 (1975).
46. K. Seshadrinathan and A. C. Bovik, "Unifying analysis of full reference image quality assessment," *IEEE Trans. Image Process.* pp. 1200–1203 (2008).

Ulrich Engelke received the Dipl.-Ing degree in electrical engineering from RWTH Aachen University, Germany, in 2004. Since 2005 he has been pursuing a PhD degree at the Blekinge Institute of Technology, Karlskrona, Sweden. His PhD studies are partly funded by a five-year scholarship awarded through the Royal Institute of Technology (KTH), Stockholm, Sweden. He is actively involved in the Joint Effort Group (JEG) of the Video Quality Experts Group (VQEG), contributing to the design and analysis of subjective experiment methodologies. He is further participating in the European FP6 Network of Excellence EuroFGI, contributing to work packages on user-perceived quality of service and quality of experience. In 2008, he was a research fellow at the University of Western Sydney, Australia, and in 2003, he conducted an internship at the Western Australian Telecommunications Research Institute (WATRI), Perth, Australia. His research interests include perceptual image and video quality assessment, visual attention modeling, machine learning techniques, and mobile multimedia communications.

Hans-Jürgen Zepernick received the Dipl.-Ing degree in electrical engineering from the University of Siegen, Germany, in 1987, and the Dr.-Ing. degree from the University of Hagen, Germany, in 1994. He is currently a professor and the chair of radio communications at the School of Engineering of the Blekinge Institute of Technology (BTH), Ronneby, Sweden. Prior to this appointment at BTH in October 2004, he held the positions of professor of wireless communications at Curtin University of Technology, Perth, Australia; Deputy Director of the Australian Telecommunications Research Institute (ATRI); and Associate Director of the Australian Telecommunications Cooperative Research Centre (ATCRC). He is an author or coauthor of some 120 technical papers and five book chapters in the areas of wireless and radio communications. He is also the lead author of the textbook entitled *Pseudo Random Signal Processing: Theory and Application* (Wiley, 2005). His research interests include radio channel modeling, radio transmission technologies, radio access techniques, mobile multimedia communications, and perceptual quality assessment of mobile multimedia services.