Visual Attention in Objective Image Quality Assessment: Based on Eye-Tracking Data

Hantao Liu, *Member, IEEE,* and Ingrid Heynderickx

***Abstract*—Since the human visual system (HVS) is the ultimate assessor of image quality, current research on the design of objective image quality metrics tends to include an important feature of the HVS, namely, visual attention. Different metrics for image quality prediction have been extended with a computational model of visual attention, but the resulting gain in reliability of the metrics so far was variable. To better understand the basic added value of including visual attention in the design of objective metrics, we used measured data of visual attention. To this end, we performed two eye-tracking experiments: one with a free-looking task and one with a quality assessment task. In the first experiment, 20 observers looked freely to 29 unimpaired original images, yielding us so-called natural scene saliency (NSS). In the second experiment, 20 different observers assessed the quality of distorted versions of the original images. The resulting saliency maps showed some differences with the NSS, and therefore, we applied both types of saliency to four different objective metrics predicting the quality of JPEG compressed images. For both types of saliency the performance gain of the metrics improved, but to a larger extent when adding the NSS. As a consequence, we further integrated NSS in several state-of-the-art quality metrics, including three full-reference metrics and two no-reference metrics, and evaluated their prediction performance for a larger set of distortions. By doing so, we evaluated whether and to what extent the addition of NSS is beneficial to objective quality prediction in general terms. In addition, we address some practical issues in the design of an attention-based metric. The eye-tracking data are made available to the research community [1].**

***Index Terms*—Eye tracking, image quality assessment, objective metric, saliency map, visual attention.**

I. Introduction

IMAGE QUALITY metrics are already integrated in a broad range of visual communication systems, e.g., for the optimization of digital imaging systems, the benchmarking of image and video coding algorithms, and the quality monitoring and control in displays [2]. These so-called objective metrics have the aim to automatically quantify the perceived image quality, and so, to serve eventually as an alternative for expensive quality evaluation by human observers. They range from dedicated metrics that measure a specific image distortion to general metrics that assess the overall perceived quality. Both the dedicated and general metrics can be classified into full-reference (FR), reduced-reference (RR), and no-reference (NR) metrics, depending on to what extent they use the original, non-degraded image or video as a reference. FR metrics are based on measuring the similarity between the distorted image and its original version. In real-world applications, where the original is not available, RR and NR metrics are used. RR metrics make use of features extracted from the original, while NR metrics attempt to assess the overall quality or some aspect of it without the use of the original.

Since the human visual system (HVS) is the ultimate assessor of image quality, it is highly desirable to have objective metrics that predict image or video quality consistent with what humans perceive [2]. Traditional FR metrics, such as the mean squared error (MSE) or the peak signal-to-noise ratio (PSNR), are simple, since they are purely defined on a pixel-by-pixel difference between the distorted and the original image, but, they are also known for their poor correlation with perceived quality [3]. Therefore, a considerable amount of research is devoted to the development of more reliable objective metrics taking characteristics of the HVS into account.

Some meaningful progress in the design of HVS-based objective metrics is reported in the literature [4]–[18]. In these studies, lower level aspects of the HVS, such as contrast sensitivity, luminance masking, and texture masking, are successfully modeled and integrated in various metrics. The basic idea behind the metrics in [4]–[7] is to decompose the image signal into channels of various frequencies and orientations in order to reflect human vision at the neural cell level. Classical HVS models, such as the contrast sensitivity function per channel, and interactions between the channels to simulate masking, are then implemented. These metrics are claimed to be perceptually more meaningful than MSE or PSNR. In [8]–[13], metrics are designed to explicitly quantify the annoyance of various compression artifacts. In this research, properties of the HVS are combined with the specific physical characteristics of the artifacts to estimate their supra-threshold visibility to the human eye. The added value of including HVS aspects in these metrics is validated with psychovisual experiments. Instead of simulating the functional components of the HVS, the metrics in [14]–[18] are rather based on the overall functionality of the HVS, e.g., by assuming that the HVS separates structural information from nonstructural information in the scene [14]. These metrics are able to

Manuscript received August 15, 2010; revised December 3, 2010; accepted February 3, 2011. Date of publication March 28, 2011; date of current version July 7, 2011. This paper was recommended by Associate Editor F. Lavagetto.

H. Liu is with the Department of Mediamatics, Delft University of Technology, Delft 2628 CD, The Netherlands (e-mail: hantao.liu@tudelft.nl).

I. Heynderickx is with the Group Visual Experiences, Philips Research Laboratories, Eindhoven 5656 AE, The Netherlands, and also with the Department of Mediamatics, Delft University of Technology, Delft 2628 CD, The Netherlands (e-mail: ingrid.heynderickx@philips.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2011.2133770

successfully predict image quality in close agreement with human judgments.

In recent years, researchers tend to include higher level aspects of the HVS, such as visual attention, in objective metrics. Limited progress has been made in this research area, mainly due to the fact that the mechanism of attention for image quality judgment is not fully understood yet, and also due to the difficulties of precisely modeling visual attention. Current research mostly incorporates visual attention into the objective metrics in an *ad hoc* way, based on optimizing the performance increase in predicting perceived quality. For example, studies in [19]–[23] are based on the assumption that a distortion occurring in an area that gets the viewer's attention is more annoying than in any other area, and they attempt to weight local distortions with local saliency, a process referred to as "visual importance pooling." The essential concept behind this approach is that the natural scene saliency (i.e., saliency driven by the original image content, and referred to as NSS) and the image distortions are taken into account separately, and they are combined to determine the overall quality score. In such a scenario, a variety of computational attention models are implemented in different metrics, resulting in a performance gain as reported in [19]–[23]. As such, this approach appears to be a viable way of including visual attention in objective metrics.

There are, however, several concerns related to the development of attention-based objective quality metrics. First of all, most research published so far in the literature employs an existing attention model to specifically optimize a targeted objective metric. Computational attention models are available, e.g., in [24] and [25], but they are either designed or chosen for a specific domain, and therefore, not necessarily generally applicable. Moreover, the accuracy of these models in predicting human visual attention is not always completely proved yet, especially not in the domain of image quality assessment. Therefore, the question arises whether an attention model successfully embedded in one particular metric is also able to enhance the performance of other metrics, and even if so, whether the gain by adding this attention model to a specific metric is comparable to the gain that can be obtained with alternative metrics. Second, it is well known that eye movements depend on the task assigned to the observer [26]. Hence, whether NSS or saliency during image quality assessment should be included in the design of objective quality metrics is still insufficiently studied. It is, e.g., not known yet whether the difference between both types of saliency is sufficiently large to actually affect the performance gain for the objective quality metrics. Third, since computational efficiency becomes a significant issue when applying an objective metric in real-time processing, the measured gain in metric performance should be balanced against the additional costs needed for the rather complex attention modeling. This implies that before implementing an attention-based metric, it is worthwhile to know exactly whether and to what extent including visual attention can improve existing objective quality metrics. Finally, studies combining visual attention and image distortions in a perceptually meaningful way are still limited, and hardly discuss a generalized strategy for combining distortion visibility and saliency.

Obviously, investigating the aspects mentioned above heavily relies on the reliability of the visual attention data used. Since recording eye movements is so far the most reliable means for studying human visual attention [26], it is highly desirable to use these "ground truth" visual attention data for the evaluation of the added value of attention in objective quality metrics. This idea is recently exploited in [27], in which the data of an eye-tracking experiment are integrated in the peak signal-to-noise ratio and structural similarity (SSIM) [14] metric. The results obtained in [27], however, are inconsistent with those found in [19]–[23], i.e., no clear improvement is found in the metric performance when weighting the local distortions with local saliency. It should, however, be noted that the eye-tracking data of [27] were collected during image quality assessment with the double stimulus impairment scale protocol [28]. This implies that each observer saw an unimpaired reference and its impaired version several times during the experiment. As a consequence, the observer might have learnt where to look for the artifacts, and thus, the recorded eye-tracking data on the impaired images may have been more affected by the image distortions than by the natural scene content. Then, simply adding these eye-tracking data to a quality metric may overweight the distraction power of the distortions compared to the NSS, and this may explain differences in the conclusions between [19]–[23] and [27] . To evaluate these assumptions, more data on whether to include NSS or saliency during scoring in the design of an attention-based metric is needed. This issue is addressed in [29] and [30], and the results show a trend of a larger improvement in predictability of the objective metrics when using eye-tracking data obtained during freely looking to unimpaired images. It should, however, be kept in mind that the study reported in [29] and [30] only made use of a limited number of human subjects (five participants looked freely to the images, while two scored the images). Nonetheless, the observed trend is in line with research recently published in [31], showing that adding "ground truth" NSS [in this case obtained by asking human observers to select the region-of-interest (ROI) in reference images] significantly improves the performance of metrics that predict the perceived quality of images that are wirelessly transmitted. Artifacts in these images are typically clustered in certain areas of the image. In such a specific scenario, using NSS is more practical since it can be transmitted as side information through the wireless communication channel. As such, the metric can make use of ROI versus background (BG) segmentation at the receiver end in real-time.

To better understand the added value of including visual attention in the design of objective metrics, we start from eye-tracking data obtained during free looking and during scoring image quality, as explained in Section II. Both types of saliency are then added to several objective quality metrics well-known in literature. The corresponding results are discussed in Section III, and reveal that although both types of saliency are beneficial for objective quality prediction, NSS tends to improve the metrics' performance more. As a consequence, we integrate, as discussed in Section IV, NSS in three full-reference metrics and two NR metrics with the aim to provide more accurate quantitative evidence on whether

and to what extent visual attention can be beneficial for objective quality prediction. We also discuss some important issues of applying NSS in the design of an attention-based metric. Moreover, we have made the eye-tracking data publicly available [1] to facilitate future research in image quality assessment.

II. Eye-Tracking Experiments

It is generally agreed that under normal circumstances human eye movements are tightly coupled to visual attention [32]–[34]. Therefore, we performed eye-tracking experiments to obtain "ground truth" visual attention data. Actually, two eye-tracking experiments were conducted. In the first experiment, the NSS for the 29 source images of the LIVE database [35] was collected by asking 20 observers to look freely to the images. In the second experiment, the saliency was recorded for 20 different observers, who were requested to score the quality of distorted versions of the source images.

A. Test Environment

The eye-tracking experiment was carried out in the New Experience Lab of the Delft University of Technology, Delft, The Netherlands [36]. Eye movements were recorded with an infrared video-based tracking system (iView X RED, SensoMotoric Instruments). It had a sampling rate of 50 Hz, a spatial resolution of 0.1°, and a gaze position accuracy of 0.5°−1.0°. Since the system could compensate for head movements within a certain range, a chin rest was sufficient to reduce head movements and ensure a constant viewing distance of 70 cm. The stimuli were displayed on a 19-in cathode ray tube monitor with a resolution of 1024 × 768 pixels and an active screen area of 365 × 275 mm. Forty students, being 24 males and 16 females, inexperienced with eye-tracking recordings, were recruited as participants. They were assigned to two groups of equal size (Groups A and B), each with 12 males and 8 females. Each session (per subject) was preceded by a 3 × 3 point grid calibration of the eye-tracking equipment.

B. Experiment I: NSS

Participants of Group A were requested to look freely to the 29 source images of the LIVE database [35]. Each participant saw all stimuli in a random order. Each stimulus was shown for 10 s followed by a mid-gray screen during 3 s. The participants were requested to look at the images in a natural way ("view it as you normally would").

C. Experiment II: Saliency During Scoring

Participants of Group B were requested to score JPEG compressed versions of the source images (using MATLAB's imwrite function). To include a broad range of quality, while avoiding that the recorded saliency was biased by viewing a scene multiple times, the source images were divided into six groups (i.e., five groups of five scenes each, and one group of four scenes, indicated by "S1" to "S6"). Each group of scenes was compressed at a different level (i.e., S1 at $Q = 5$, S2 at $Q = 10$, S3 at $Q = 15$, S4 at $Q = 20$, S5 at $Q = 30$, and S6 at $Q = 40$). By doing so, each scene was viewed only once per subject, and for each subject in a different random order. The subject was requested to score the image quality for each stimulus with the single-stimulus (SS) method, i.e., in the absence of a reference [28]. A categorical scoring scale (recommended by ITU-R [28]) with the semantic terms "Excellent," "Good," "Fair," "Poor," and "Bad" was used. Each stimulus was shown for 10 s, followed by a scoring screen as illustrated in Fig. 1. The actual experiment was preceded by a training, in which the participant was instructed on the task and could familiarize himself/herself with how to use the scoring scale.

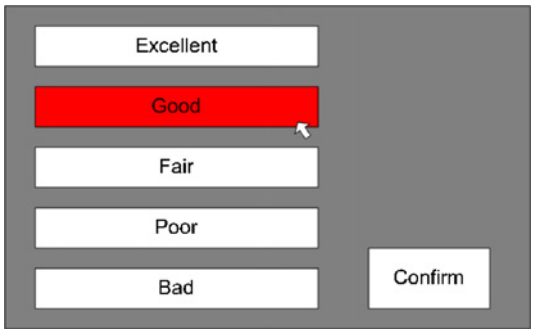


Fig. 1. Illustration of the scoring screen.

III. NSS Versus Saliency During Scoring Applied in Objective Metrics

A. Saliency Map

A saliency map representative for visual attention is usually derived from the spatial pattern of fixations in the eye tracking data [32]–[34]. To construct this map, each fixation location gives rise to a grayscale patch whose activity is Gaussian distributed. The width (σ) of the Gaussian patch approximates the size of the fovea (about 2° of visual angle). A mean saliency map (MSM) over all fixations of all subjects is then calculated as follows:

$$S_i(k,l) = \sum_{j=1}^{T} \exp\left[-\frac{(x_j-k)^2+(y_j-l)^2}{\sigma^2}\right] \quad (1)$$

where $S_i(k,l)$ indicates the saliency map for stimulus I_i of size $M \times N$ pixels (i.e., $k \in [1, M]$ and $l \in [1, N]$), (x_j, y_j) are the spatial coordinates of the jth fixation ($j = 1, \ldots, T$), T is the total number of all fixations over all subjects, and σ indicates the standard deviation of the Gaussian (i.e., $\sigma = 45$ pixels in our specific case). The intensity of the resulting saliency map is linearly normalized to the range [0, 1]. Fig. 2 illustrates as an example a MSM derived from eye-tracking data obtained in experiment I for one of the original images, and the MSM obtained in experiment II for a JPEG compressed version of the same image (the saliency maps for the entire database can be accessed in [1]).

The example illustrates typical correspondences and differences between the NSS, derived from experiment I, and the

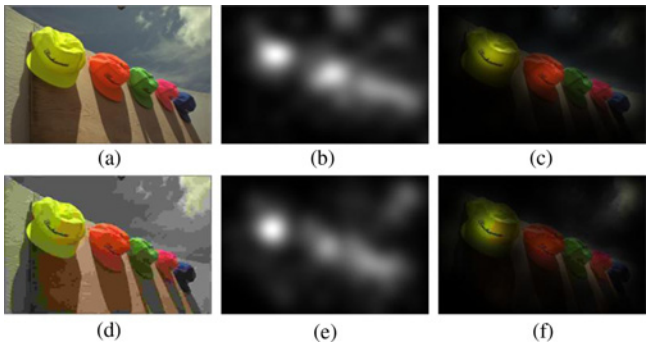


Fig. 2. Illustration of the saliency map. (a) Original image. (b) MSM of (a) derived from the eye-tracking data of experiment I. (c) Saliency map (b) superimposed on the image (a). (d) JPEG compressed image (Q = 5). (e) MSM of (d) derived from the eye-tracking data of experiment II. (f) Saliency map (e) superimposed on the image (d). Note that the darker the regions are, the lower the saliency is.

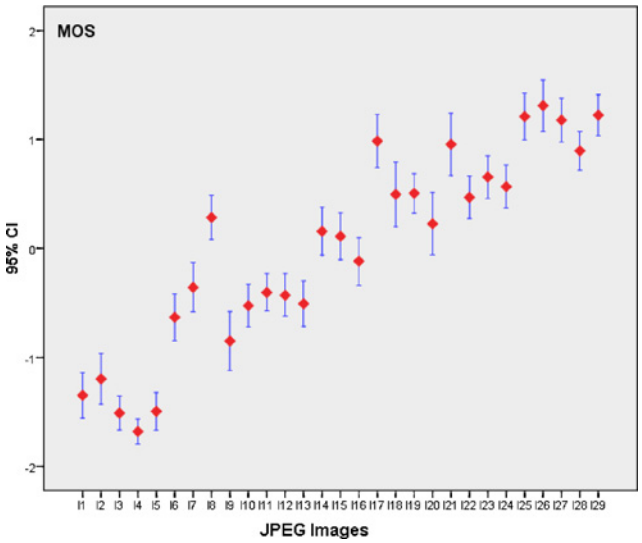


Fig. 3. MOS of the 29 JPEG images of experiment II. The error bars indicate the 95% confidence interval.

saliency during scoring, derived from experiment II. In general, the most salient regions are comparable between the NSS and the saliency during scoring, but there are some deviations for which it is worthwhile to investigate their impact on the performance of an objective metric. An extensive discussion on the differences between NSS and saliency during scoring, including aspects of the appropriate comparison method, and the impact of the experimental protocol, is outside the scope of this paper, and will be treated in a separate contribution [37].

B. Added Value of NSS and Saliency During Scoring in Objective Metrics

Based on the eye-tracking data, obtained from both our experiments, we evaluate whether and to what extent adding saliency is beneficial to the prediction performance of objective metrics. In this evaluation, we compare the performance gain obtained when adding NSS versus saliency during scoring. To this end, we use the subjective scores we obtained in experiment II, and we try to predict these scores with several well-known objective metrics, all weighted with both types of saliency.

1) *Subjective Scores:* In experiment II, 20 human subjects scored the quality of 29 JPEG distorted images. We transformed the raw quality ratings (i.e., "Excellent" = 5, "Good" = 4, "Fair" = 3, "Poor" = 2, and "Bad" = 1 as shown in Fig. 1) into numbers, and calculated the mean opinion score (MOS) as described in [13]. The resulting MOS are illustrated in Fig. 3.

2) *Objective Metrics:* The evaluation of adding saliency was performed with four objective metrics (i.e., three FR metrics and one NR metric), which are so far widely accepted in the image quality community to assess the quality of JPEG compressed images. The FR metrics are as follows.

a) *PSNR*: The peak signal-to-noise ratio simply measures the difference (i.e., MSE) between the distorted image and its original version on a pixel-by-pixel base.

b) *SSIM*: The structural similarity index [14] assumes that the HVS is highly adapted for extracting structural information from a scene, and it measures image quality based on the degradation in structural information.

c) *VIF*: The visual information fidelity [15] quantifies how much of the information present in the reference image can be extracted from the distorted image. Note that in this paper, we use the implementation of the VIF in the spatial domain (as described in [35]).

The NR metric is as follows.

a) *GBIM*: The generalized block-edge impairment metric [8] is one of the most well-known metrics to quantify blocking artifacts in discrete cosine transform (DCT) coding. It measures blockiness as an inter-pixel difference across block boundaries (i.e., referred to as block-edges) scaled with a weighting function, which addresses luminance and texture masking of the HVS.

The objective metrics mentioned above are all formulated in the spatial domain. They estimate the image distortion locally, yielding a quantitative distortion map, which provides a spatially varying quality degradation profile. As an example, Fig. 4(a) illustrates the distortion map calculated by SSIM for the JPEG compressed image of Fig. 2(d) (bit rate of 0.41 b/p). The intensity value of each pixel in the distortion map indicates the local degree of distortion, i.e., the lower the intensity, the larger the distortion is.

3) *Including Saliency:* Saliency (i.e., either NSS or saliency during scoring) is included in a metric by locally weighting the distortion map, as illustrated in Fig. 4(b) and (c) for the distortion map of SSIM weighted with NSS and saliency during scoring, respectively. Note that in the case of GBIM, the metric is calculated only around block-edges. As a result, weighting its distortion map with saliency actually gives more weight to the block-edges in the salient areas than in the non-salient areas.

Adding saliency to PSNR, SSIM, VIF, and GBIM results in eight attention-based metrics, which are referred to as WPSNR_NSS, WPSNR_SS, WSSIM_NSS, WSSIM_SS, WVIF_NSS, WVIF_SS, WGBIM_NSS, and WGBIM_SS,

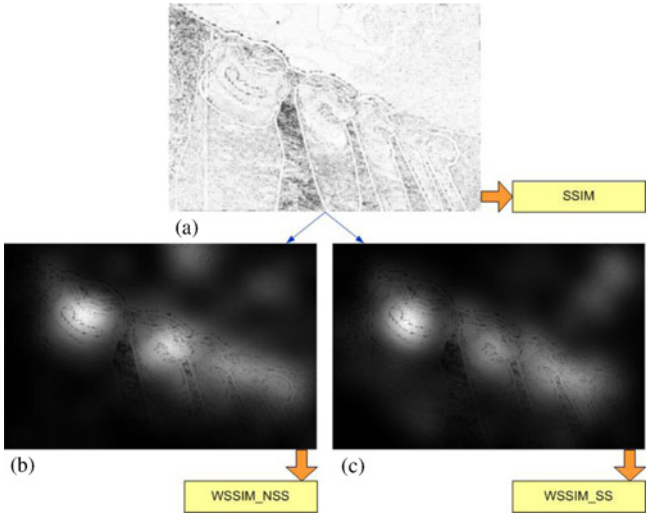
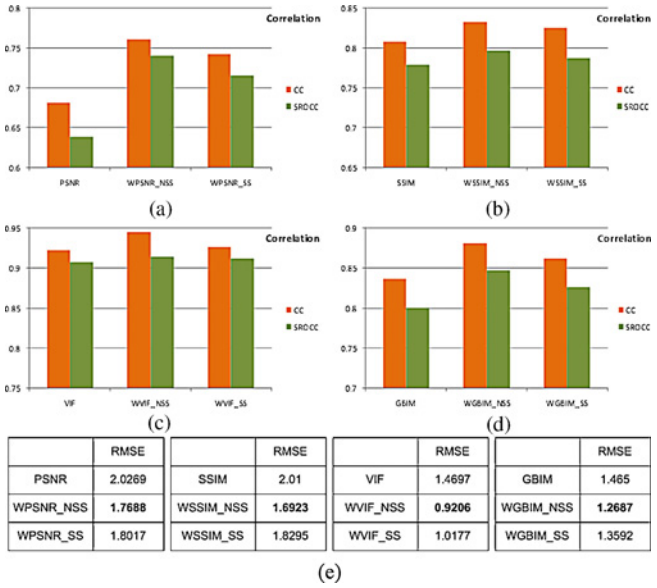


Fig. 4. Illustration of an objective metric based on saliency. (a) distortion map of SSIM calculated for the JPEG compressed image (bit rate 0.41 b/p) of Fig. 2(d). (b) Corresponding NSS superimposed on (a). (c) Corresponding saliency during scoring superimposed on (a). For the distortion map, the lower the intensity, the larger the distortion is.



	RMSE		RMSE		RMSE		RMSE
PSNR	2.0269	SSIM	2.01	VIF	1.4697	GBIM	1.465
WPSNR_NSS	**1.7688**	WSSIM_NSS	**1.6923**	WVIF_NSS	**0.9206**	WGBIM_NSS	**1.2687**
WPSNR_SS	1.8017	WSSIM_SS	1.8295	WVIF_SS	1.0177	WGBIM_SS	1.3592

(e)

Fig. 5. CCs (without nonlinear regression) of 12 metrics for the 29 JPEG images of experiment II. (a) PSNR, WPSNR_NSS, and WPSNR_SS. (b) SSIM, WSSIM_NSS, and WSSIM_SS. (c) VIF, WVIF_NSS, and WVIF_SS. (d) GBIM, WGBIM_NSS, and WGBIM_SS. The corresponding RMSE-values are given in (e).

respectively. They can be defined as follows:

$$WMetric = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}[distortion\_map(x, y) \cdot S_i(x, y)]}{\sum_{x=1}^{M}\sum_{y=1}^{N} S_i(x, y)} \quad (2)$$

where distortion_map is calculated by the metric used, S indicates the corresponding saliency map derived from the eye-tracking experiment, and WMetric denotes the resulting attention-based metric. It should be noted that the combination strategy used here is a simple weighting function similar to that in [19]–[23]. More complex combination strategies may further improve the metric's performance, as is discussed in Section IV.

4) *Experimental Results:* As prescribed by the Video Quality Experts Group [38], the performance of an objective metric is determined by its ability to predict subjective quality ratings (the MOS). This ability can be quantified by the Pearson linear correlation coefficient (CC) indicating prediction accuracy, the Spearman rank order correlation coefficient (SROCC) indicating prediction monotonicity, and the root-mean-squared error (RMSE). With respect to the latter measure, we want to note that the scores are normalized to the scale [1], [10] before the calculation of the RMSE. As suggested in [38], the metric's performance can also be evaluated with nonlinear correlations using a nonlinear mapping of the objective predictions before computing the correlation. Indeed, the image quality community is more accustomed to, e.g., a logistic function, to fit the predictions of an objective metric to the MOS. It may, e.g., account for a possible saturation effect in the quality scores at high quality. A nonlinear fitting usually yields higher CCs in absolute terms, while generally keeping the relative differences between the metrics [39]. On the contrary, without a sophisticated nonlinear fitting (often including additional parameters) the CCs cannot mask a bad performance of the metric itself, as discussed in [23]. To better visualize differences in performance, we avoid any nonlinear fitting and directly use linear correlation and RMSE between the metrics' predictions and the MOS.

The 12 metrics (i.e., PSNR, WPSNR_NSS, WPSNR_SS, SSIM, WSSIM_NSS, WSSIM_SS, VIF, WVIF_NSS, WVIF_SS, GBIM, WGBIM_NSS, and WGBIM_SS) are applied to the 29 JPEG compressed images, and the results are compared to the corresponding MOS of experiment II. Fig. 5 shows the resulting CC, SROCC, and RMSE-values, and demonstrates that the performance of all metrics enhances by including both NSS and saliency during scoring. The experimental results also tend to indicate that adding NSS to a metric yields a larger amount of performance gain than adding saliency during scoring. Adding NSS to PSNR corresponds to an increase of 8% in CC and of 10% in SROCC, and a decrease of 0.258 in the RMSE value, but adding saliency during scoring to PSNR results only in an increase of 6% in CC and of 8% in SROCC, and a decrease of 0.225 in the RMSE value. The same trend of changes in performance is consistently found for the three other metrics.

Based on the above results, we can conclude that the small difference in saliency due to scoring with respect to the NSS is nonetheless sufficient to yield a consistent difference in performance gain when including visual attention to objective metrics. The relatively lower performance gain obtained with the saliency during scoring is possibly caused by the fact that this saliency is more spread toward BG areas in the image due to the distraction power of annoying artifacts. As such, artifacts in BG areas are weighted more (in relative terms) than artifacts in salient areas, and so, this might result in an overestimation of the annoyance of distortions in the BG. Our results tend to support the assumption made in Section I for

the difference in conclusion given in [27], on the one hand, and in [19]–[23], on the other hand. When adding saliency to objective metrics, it should be the NSS, obtained when people look at a distortion-free image for the first time. The saliency or distraction power of the image distortions themselves is kind of addressed by the metric (especially, when HVS aspects, such as contrast sensitivity and masking are already included in the distortion map).

IV. Adding NSS in Objective Metrics: Based on LIVE Database

To further evaluate the added value of visual attention in objective metrics, we include the NSS obtained from our eye-tracking data in experiment I into various objective metrics available in literature, and compare the performance of these attention-based metrics to the performance of the same metrics without visual attention. To also evaluate a variety of distortion types, this validation is done for the entire LIVE database [35], which consists of 779 images distorted with JPEG compression (i.e., JPEG), JPEG2000 compression (i.e., JP2 K), white noise (i.e., WN), Gaussian blur (i.e., GBLUR), and simulated fast-fading Rayleigh occurring in (wireless) channels (i.e., FF). Per image the database also gives a difference in mean opinion score (DMOS) derived from an extensive subjective quality assessment study [40]. Based on the evaluation, we address some technical issues relevant to the application of visual attention in objective metrics. More specifically, we discuss the effect of image content and of the combination strategy.

A. Objective Metrics

For practical reasons, the objective metrics used in our validation are limited to three well-known FR metrics and two NR metrics. The FR metrics are PSNR, SSIM, and VIF, as explained in Section III. The NR metrics are GBIM (also explained in Section III) and NR perceptual blur (NRPB). The latter refers to the NRPB metric [11] based on extracting sharp edges in an image, and measuring the width of these edges.

B. Evaluation of the Overall Performance Gain

Adding NSS to the metrics mentioned above results in five attention-based metrics, which are referred to as WPSNR, WSSIM, WVIF, WGBIM, and WNRPB, respectively. The six FR metrics, i.e., PSNR, SSIM, VIF, WPSNR, WSSIM, and WVIF, are intended to assess image quality independent of distortion type, and therefore, are applied to the entire LIVE database [35]. The metrics GBIM and WGBIM are designed specifically for block-based DCT compression, and are applied to the JPEG#1 and JPEG#2 subsets of the LIVE database. The metrics NRPB and WNRPB are designed to quantify blur in images, and they are applied to the GBLUR subset of the LIVE database.

Figs. 6 and 7 give the corresponding CCs and RMSE values. The overall gain (averaged over artifacts where appropriate) of an attention-based metric over its corresponding metric without NSS is summarized in Tables I and II. Both figures and tables demonstrate that there is indeed a gain in performance when including visual attention in the objective metrics PSNR, SSIM, VIF, GBIM, and NRPB, independent of the metric used and of the image distortion type tested. The actual amount of performance gain, however, depends on the metric and on the distortion type. A promising performance gain (expressed in terms of CC) is found for the subset of the LIVE database distorted by GBLUR: the gain of WPSNR over PSNR is 2%, of WSSIM over SSIM is 7%, of WVIF over VIF is 2%, and of WNRPB over NRPB is 5%. The amount of performance gain, however, is relatively small for the subset of the LIVE database distorted by WN: the gain (again in terms of CC) of WPSNR over PSNR is 0.01%, of WSSIM over SSIM is 1%, and of WVIF over VIF is 1%. Differences in performance may be attributed to two possible causes: 1) the performance of a metric (i.e., without NSS) varies with the distortion type, and as such it is more difficult to obtain a significant increase in performance by adding NSS when a metric already has a high prediction performance for a given type of distortion, and 2) in the specific case of images distorted by GBLUR, some metrics might confuse unintended (Gaussian) blur with intended blur in the BG to increase the field of depth (i.e., a high-quality foreground object with an intentionally blurred BG). Adding NSS reduces the importance of blur in the BG, and as such might improve the overall prediction performance of a metric.

C. Statistical Significance

In order to check whether the numerical difference in performance between a metric with NSS and the same metric without NSS is statistically significant, we performed some hypothesis testing to provide statistical soundness on the conclusion of superiority of the attention-based metrics. As suggested in [38], the test is based on the residuals between the DMOS and the quality predicted by the metric (hereafter, referred to as M-DMOS residuals). Before being able to do a parametric test, we evaluated the assumption of normality of the M-DMOS residuals. A simple kurtosis-based criterion (as used in [40]) was used for normality; if the residuals had a kurtosis between 2 and 4, they were assumed to be normally distributed, and the difference between the two sets of M-DMOS residuals could be tested with a parametric test. The results of the test for normality are summarized in Table III, and indicate that in most cases the residuals are normally distributed. Considering that most parametric tests are not too sensitive to deviations from normality, we decided to test statistical significance for the performance improvement of NSS-based metrics with a parametric test for all combinations of objective metrics with distortion types. In our particular case, the two sets of residuals being compared are dependent samples: one is from the metric itself and one is from the same metric after adding the NSS. Therefore, a paired-sample t-test [41] is used instead of the F-test, as suggested in [38], since the latter one assumes that the two samples being compared are independent. The paired-sample t-test starts from the null hypothesis stating that the residuals of one metric are statistically indistinguishable (with 95% confidence) from the residuals of that same metric with NSS. The results of this t-test are given in Table IV for all metrics and distortion types separately. This table illustrates that in most cases the improvement in prediction performance

TABLE I

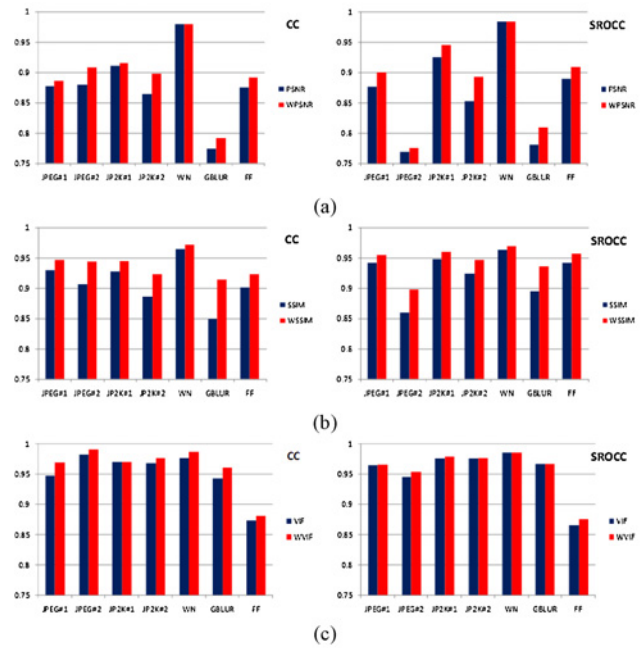
PERFORMANCE OF PSNR, WPSNR, SSIM, WSSIM, VIF, AND WVIF AVERAGED OVER ALL DISTORTION TYPES FOR THE IMAGES OF THE LIVE DATABASE [35]

	CC	SROCC	RMSE		CC	SROCC	RMSE		CC	SROCC	RMSE
PSNR	0.88	0.87	1.09	SSIM	0.91	0.92	0.86	VIF	0.95	0.955	0.70
WSNR	0.90	0.90	0.99	WSSIM	0.94	0.95	0.74	VWIF	0.96	0.958	0.62
Δ	ΔP = 2%	ΔS = 3%	ΔR = 0.1	Δ	ΔP = 3%	ΔS = 3%	ΔR = 0.12	Δ	ΔP = 1%	ΔS = 0.3%	ΔR = 0.08

TABLE II

PERFORMANCE OF GBIM AND WGBIM FOR THE SUBSETS JPEG#1 AND JPEG#2, AND PERFORMANCE OF NRPB AND WNRPB FOR THE SUBSET GBLUR OF THE LIVE DATABASE [35]

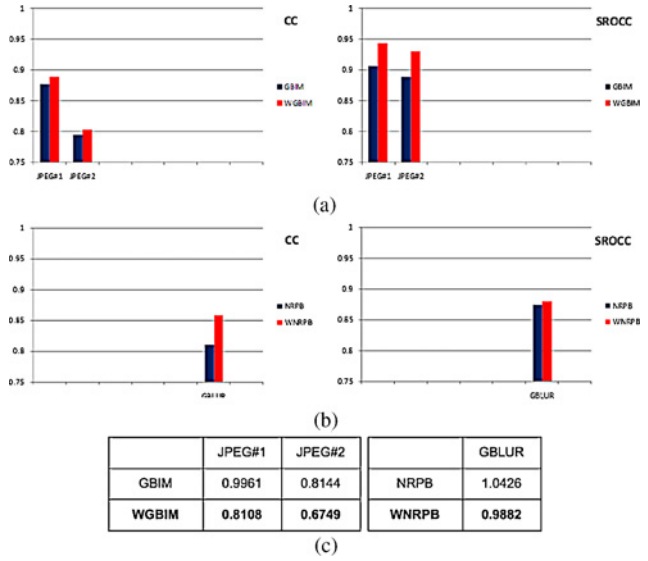
	CC	SROCC	RMSE		CC	SROCC	RMSE
GBIM	0.83	0.90	0.91	NRPB	0.81	0.87	1.04
WGBIM	0.84	0.94	0.74	WNRPB	0.86	0.88	0.99
Δ	ΔP = 1%	ΔS = 4%	ΔR = 0.17	Δ	ΔP = 5%	ΔS = 1%	ΔR = 0.05



	JPEG#1	JPEG#2	JP2K#1	JP2K#2	WN	GBLUR	FF
PSNR	1.0586	1.2908	0.9926	1.3382	0.4422	1.3580	1.1585
WPSNR	**0.9853**	**1.1168**	**0.9050**	**1.1341**	**0.4383**	**1.2881**	**1.0578**
SSIM	0.7820	0.7540	0.8383	1.0441	0.6290	1.0666	0.9046
WSSIM	**0.6966**	**0.6317**	**0.7307**	**0.8479**	**0.6000**	**0.8282**	**0.8406**
VIF	0.6881	0.4881	0.5510	0.6563	0.5693	0.6762	1.2371
WVIF	**0.5882**	**0.4069**	**0.4964**	**0.5853**	**0.4364**	**0.6350**	**1.1933**

(d)

Fig. 6. CCs (without nonlinear regression) of six FR metrics for images distorted by JPEG#1, JPEG#2, JPEG2000#1, JPEG2000#2, WN, GBLUR, and FF, respectively. (a) PSNR and WPSNR. (b) SSIM and WSSIM. (c) VIF and WVIF. The corresponding RMSE-values are given in (d). Note that the data used are taken from the LIVE database [35].



	JPEG#1	JPEG#2
GBIM	0.9961	0.8144
WGBIM	**0.8108**	**0.6749**

	GBLUR
NRPB	1.0426
WNRPB	**0.9882**

(c)

Fig. 7. CCs (without nonlinear regression) of four NR metrics. (a) GBIM and WGBIM for JPEG#1 and JPEG#2. (b) NRPB and WNRPB for GBLUR. The corresponding RMSE-values are given in (c). Note that the data used are taken from the LIVE database [35].

by adding NSS to an objective metric is statistically significant. The improvement reported in Section IV-B is not statistically significant only in three combinations of metrics applied to a given distortion type (with only 29 stimuli).

It should, however, be noted that statistical significance testing is not straightforward, and the conclusions drawn from it largely depend, e.g., on the number of sample points, on the selection of the confidence criterion, and on the assumption of normality of the residuals. These issues are extensively discussed in [40].

D. Evaluation of the Influence of Image Content

The distribution of saliency over an image largely depends on its content, and, therefore, it makes sense to also study whether the added value of including visual attention to objective metrics is content dependent. The effect of content on NSS is quantified by calculating per image the correlation between the MSM obtained from experiment I and each individual saliency map (ISM) (derived from the fixations of an individual subject). The correlation between two saliency maps (i.e., SM_A and SM_B) is often measured by the coefficient (ρ), as employed in [32]. It is defined with its value ranging

TABLE III
NORMALITY OF THE M-DMOS RESIDUALS

	JPEG#1	JPEG#2	JP2K#1	JP2K#2	WN	GBLUR	FF
PSNR	1	1	1	1	1	1	1
WPSNR	1	1	1	1	1	1	1
SSIM	1	1	1	1	1	0	1
WSSIM	1	1	1	1	1	0	1
VIF	1	1	1	1	1	1	1
WVIF	1	1	1	1	1	1	1
GBIM	1	0					
WGBIM	1	0					
NRPB						1	
WNRPB						1	

"1" means that the residuals can be assumed to have a normal distribution since the kurtosis lies between 2 and 4.

TABLE IV
RESULTS OF t-TEST BASED ON M-DMOS RESIDUALS

	JPEG#1	JPEG#2	JP2K#1	JP2K#2	WN	GBLUR	FF
PSNR and WPSNR	1	1	1	–	1	1	–
SSIM and WSSIM	1	1	1	1	1	1	1
VIF and WVIF	1	1	1	1	1	1	1
GBIM and WGBIM	1	–					
NRPB and WNRPB						1	

"1" means that the attention-based metric is statistically significantly better that the metric without NSS, and "–" means that the difference is not statistically significant.

$[-1, 1]$ as follows:

$$\rho = \frac{\sum_{n=1}^{M}(SM_A(n) - \mu_A)(SM_B(n) - \mu_B)}{\sqrt{\sum_{n=1}^{M}(SM_A(n) - \mu_A)^2 \sum_{n=1}^{M}(SM_B(n) - \mu_B)^2}} \quad (3)$$

where μ_A and μ_B are the mean values of the SM_A and SM_B, respectively. M is the total number of pixels in both maps. A higher value of ρ indicates a larger similarity between the two saliency maps. Fig. 8 gives the ρ-values between the MSM and the ISM averaged over all subjects. This averaged ρ-value strongly varies over the different natural scenes, with the highest value of ρ for "scene25" ($\rho = 0.7549$) and the lowest value of ρ for "scene3" ($\rho = 0.4521$). This averaged ρ-value quantifies the variation in eye-tracking behavior among human subjects when viewing a SS. A large value of the ρ averaged over all subjects indicates a small variation in saliency among subjects, while a small value of ρ indicates that the saliency is widely spread among subjects. Fig. 9 presents the images with the three smallest values of the averaged ρ (i.e., "set_low") in Fig. 8. These images clearly lack highly salient features, and their corresponding MSM includes fixations distributed all over the image. Fig. 10 shows the three images, with the largest value of the averaged ρ (i.e., "set_high") in Fig. 8. These images generally contain a few salient features, such as the human face in the images "statue" and "studentsculpture" and the billboard in the image "cemetry." For these images, the saliency converges around these features in the MSM. The difference in saliency between both sets of images is apparently driven by the image content.

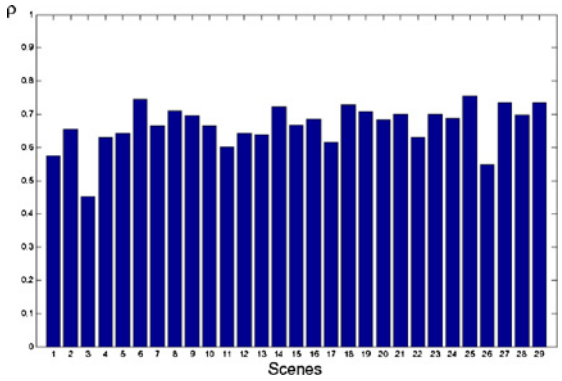


Fig. 8. CC (ρ) between the MSM and the ISM averaged over all subjects per scene.

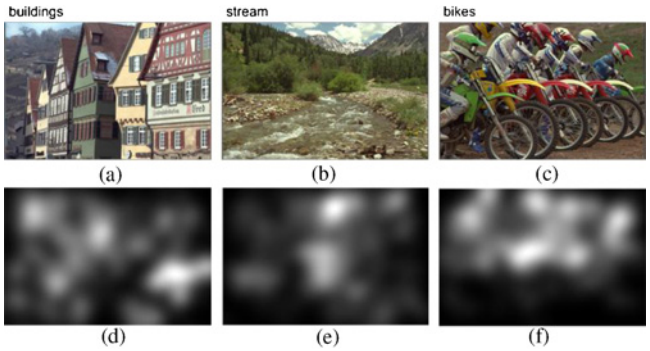


Fig. 9. Illustration of the three images with the smallest correspondence in saliency between subjects (i.e., smallest value of averaged ρ in Fig. 8). (a) $\rho = 0.4521$. (b) $\rho = 0.5485$. (c) $\rho = 0.5963$. (d) MSM-buildings. (e) MSM-streams. (f) MSM-bikes.

To evaluate the content dependency in the performance gain when adding saliency to objective metrics, we repeated the experiment in Section IV-B once for the source images of "set_low," and once for the source images of "set_high." The former set contained 20 stimuli with JPEG compression, 17 stimuli with JPEG2000 compression, 15 stimuli with WN, 15 GBLUR stimuli, and 15 stimuli with FF artifacts, while the latter set consisted of 18 stimuli with JPEG compression, 17 stimuli with JPEG2000 compression, 15 stimuli with WN, 15 GBLUR stimuli, and 15 stimuli with FF artifacts. Fig. 11 illustrates the comparison in performance gain (i.e., quantified by the Pearson CC) between a metric and its NSS-weighted version for the "set_low" and "set_high" images separately. In general, it shows the consistent trend that including saliency results in a larger performance gain in the objective metrics for the images of "set_high" than for the images of "set_low"; more particularly, for the images of "set_low," the performance gain when adding saliency is actually non-existing. The gain of WPSNR over PSNR corresponds to an average increase in the Pearson CC (over all distortion types of the LIVE database) from 0.942 to 0.943 for the "set_low" images (i.e., 0.1%), and from 0.882 to 0.910 for the "set_high" images (i.e., 2.8%). The gain of WSSIM over SSIM is 0 (from 0.976 to 0.976) for the "set_low" images and 3.1% (from 0.934 to 0.965) for the "set_high" images. The gain of WVIF over VIF is 0 (from 0.958 to 0.958) for the "set_low" images and

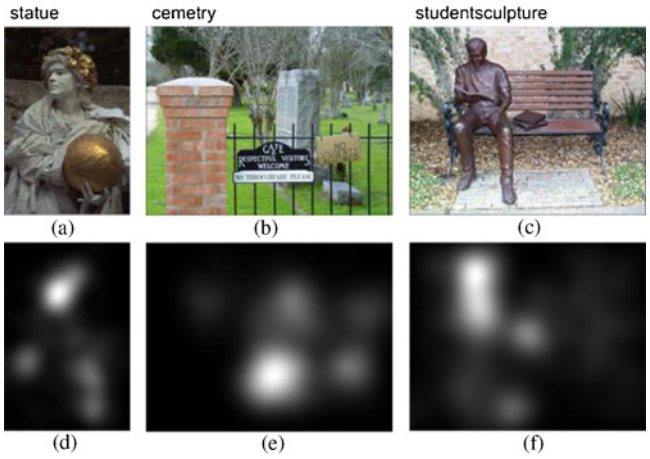


Fig. 10. Illustration of the three images with the largest correspondence in saliency between subjects (i.e., largest values of the averaged ρ in Fig. 8). (a) ρ = 0.7549. (b) ρ = 0.7444. (c) ρ = 0.7344. (d) MSM-statue. (e) MSM-cemetry. (f) MSM-studentsculpture.

1.6% (from 0.966 to 0.982) for the "set_high" images. The gain of WGBIM over GBIM is 1.6% (from 0.929 to 0.945) for the "set_low" images and 7.7% (from 0.789 to 0.866) for the "set_high" images. There is, however, one exception to this trend, namely, for the metrics WNRPB and NRPB. As shown in Fig. 11(e), adding saliency degrades the performance of NRPB for the images of "set_high." This may be due to the specific design of the blur metric, which is based on measuring the width of extracted strong edges. Including the saliency of Fig. 10 to the NRPB metric with a linear weighting combination strategy runs the risk of eliminating some very obvious edges in the calculation of blur, and may consequently affect the accuracy of the metric.

In summary, our findings suggest that the performance gain in an objective metric when applying saliency depends on the image content as well as on the specific metric design.

E. Evaluation of the Influence of Combination Strategy

So far, saliency was added to the objective metrics based on a linear weighting combination strategy. This method is simple and intuitive, and has been widely adopted to pool local distortions of an image with saliency [19]–[23]. Our results of Sections III and IV demonstrate the general effectiveness of using the linear combination strategy. This strategy, however, has limitations in dealing with certain distortions in more demanding conditions [42]. Fig. 12 illustrates an image JPEG compressed at a bit rate of 0.43 b/p, and its corresponding NSS obtained from our eye-tracking data. Due to texture and luminance masking in the HVS [10], this image exhibits imperceptible blocking artifacts in the more salient areas (e.g., the foreground of the white tower), and relatively annoying blocking artifacts in the less salient areas (e.g., the BG of the sky). In such a case, combining the distortion and saliency map with a linear combination strategy intrinsically underestimates the annoyance of the artifacts in the BG, and their impact on the quality judgment.

To quantify the effect of linearly adding saliency in an objective metric for the quality prediction of demanding images, a subset of nine images was selected from the LIVE database. The images "img{9, 37, 44, 47, 63, 69, 89, 92, 105}" of the subset JPEG#1 typically represent the type of JPEG compressed images with the artifacts in the more salient areas locally masked by the content, and with clearly visible artifacts in the less salient areas. The blockiness metrics, GBIM and WGBIM are applied to this sub-selection of the database. As illustrated in Fig. 13, WGBIM fails in accurately predicting the subjective quality ratings for this subset of demanding images, mainly due to the inappropriate integration of saliency in the blockiness metric (i.e., the gain of WGBIM over GBIM in CC is −59%). Hence, the overall gain in CC of WGBIM over GBIM (i.e., 1%) for the entire LIVE database of JPEG compressed images is explained by the fact that most of the

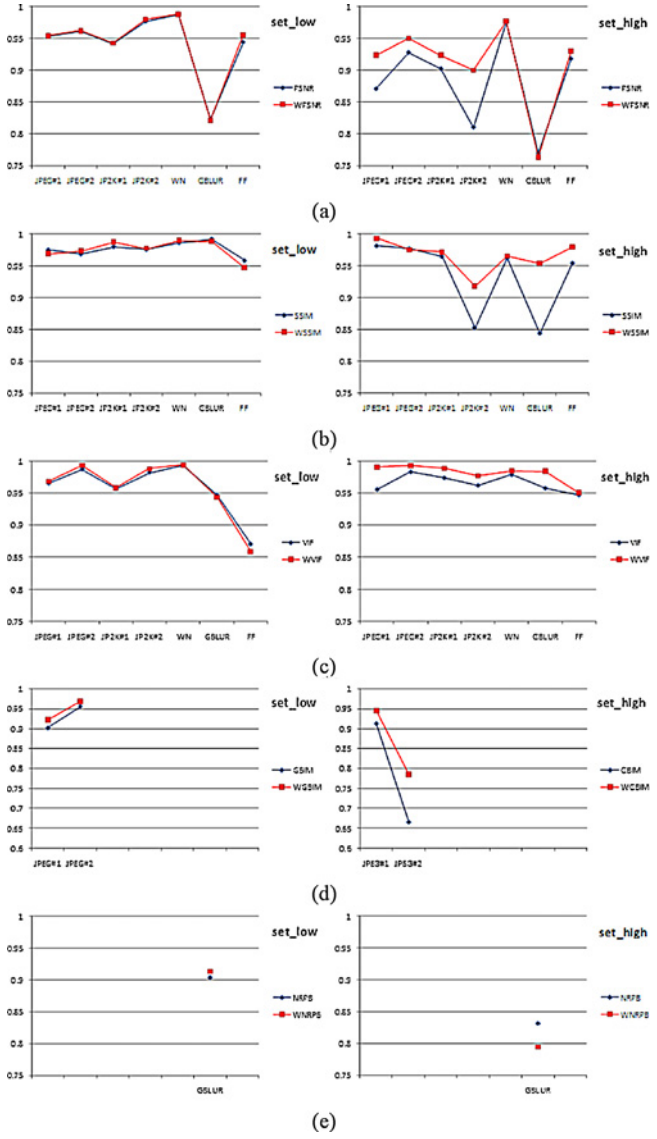
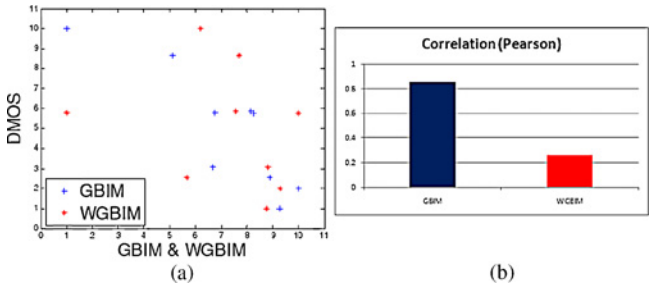


Fig. 11. Comparison in performance gain when adding saliency (quantified by the Pearson CC) between images of "set_low" (distorted images extracted from the LIVE database [35] based on the source images of Fig. 9) and images of "set_high" (distorted images extracted from the LIVE database [35] based on the source images of Fig. 10). (a) PSNR versus WPSNR. (b) SSIM versus WSSIM. (c) VIF versus WVIF. (d) GBIM versus WGBIM. (e) NRPB versus WNRPB.

(a) (b)

Fig. 12. (a) Image JPEG compressed at a bit rate of 0.43 b/p, and (b) its corresponding NSS obtained from our eye-tracking data.



(a) (b)

Fig. 13. Performance of the blockiness metrics GBIM and WGBIM in predicting the subjective quality rating of a subset of demanding images (i.e., img{9, 37, 44, 47, 63, 69, 89, 92, 105}) selected from the LIVE database JPEG#1 [35]. (a) Scatter plot of DMOS versus GBIM and WGBIM. (b) CCs (without nonlinear regression) of GBIM and WGBIM.

images in this database exist of one of the following types: 1) images having visible artifacts uniformly distributed over the entire image, and 2) images having the artifacts masked by the content in the less salient areas, but showing visible artifacts in the more salient areas. Obviously, for these two types of images, adding saliency with a linear combination strategy is reasonable.

So, these findings indicate that a linear combination strategy is not necessarily appropriate for adding saliency in objective metrics. Hence, from a point of view of metric optimization, it is worthwhile to investigate adaptive combination strategies as, e.g., discussed in [23] and [42].

V. Discussion

In this paper, we evaluate the intrinsic gain in prediction accuracy that can be obtained by introducing visual attention in objective quality metrics. This evaluation is performed for a diverse, though limited set of images, and mainly for distortions that affect the images globally. The results we obtained show that there is added value in weighting pixel-based distortion maps with local saliency. The amount of added value is bigger when extending the objective metrics with NSS than with saliency recorded while the viewers assess the quality of the images. The actual gain in performance accuracy is highly dependent on the image content, on the distortion type, and on the objective metric itself. Images with a clear ROI demonstrate a bigger gain as compared to images in which the NSS is spread over the whole image. In addition, the gain is small for objective metrics that already show a high correlation with perceived quality for a given distortion type.

Although showing clear results, the study reported here has some limitations. First, as mentioned above, the set of images used has a fair size, but could be extended in order to investigate the effect of image content on the gain in prediction accuracy in a more systematic way. Second, most images are degraded with distortions that affect the image quality globally, i.e., the artifacts are uniformly distributed over the entire image. In specific applications, such as in wireless imaging, artifacts may occur localized, i.e., only at some random, but limited location in the image. Although we did not investigate this type of distortions specifically, we expect that introducing visual saliency in quality prediction metrics for this type of distortions is still beneficial. At least, results reported in [31] support this hypothesis. Finally, the gain in prediction accuracy claimed in this paper is based on eye-tracking recordings. These recordings intrinsically have some inaccuracy, which may limit the overall reliability of our conclusions. We have shown, however, that recorded saliency data are highly consistent when using well-calibrated equipment and a well-defined protocol; the consistency is even shown for data collected in various laboratories [43]. Using eye-tracking data, of course, is unrealistic for real-time applications. Hence, a visual attention model will be needed in the actual implementation of an objective metric. Since the reliability of most visual attention models is still limited, we expect that the actual gain in prediction accuracy that can be obtained in a real-time application is lower than what we showed here, at least with the current soundness of visual attention models. In the coming years, the soundness of visual attention models may improve, but most probably at the expense of their computational cost.

Given the fact that the added value of having NSS weighted objective quality metrics depends on the image content, distortion type, and objective metric, an adaptive approach might be desirable in real-time applications to limit the overall computational cost. In such an approach, the performance of an objective metric needed in the video chain can be optimized offline, i.e., for each metric the added value of incorporating saliency can be estimated from its general prediction accuracy. For those metrics that contain saliency in their extended version a simple visual attention model can be used to determine the size of the ROI in the image. Only when the ROI is limited in size, the extended version of the metric is needed. Otherwise, the metric without saliency model can be applied at sufficient accuracy.

VI. Conclusion

In this paper, we investigated the added value of visual attention in the design of objective metrics. Instead of using a computational model for visual attention, we conducted eye-

tracking experiments to obtain "ground truth" visual attention data, thus making the results independent of the reliability of an attention model. Actually, two eye-tracking experiments were performed: one in which the participants looked freely to undistorted images, and a second one in which different participants were asked to score the quality of a JPEG compressed version of the images. The resulting eye-tracking data indicated that there is some deviation between the NSS and saliency during scoring.

Adding either type of saliency to an objective metric improved its performance in predicting perceived image quality. However, we also found a tendency that adding NSS to a metric yields a larger amount of gain in the performance. Based on this evidence, the data of NSS were further integrated in several objective metrics available in literature, including three FR metrics and two NR metrics. This evaluation showed that there is indeed a gain in the performance for all these metrics when linearly weighting the local distortion map of the metrics with the NSS. The extent of the performance gain tends to depend on the specific objective metric and the image content. But our findings also illustrated that for some image content and for some distortion types, the linear combination strategy is insufficient and adaptive strategies are needed. Current and future research includes modeling saliency for real-time quality assessment, and integrating this saliency in objective metrics in a perceptually even more meaningful way.

REFERENCES

[1] H. Liu and I. Heynderickx. (2010). *TUD Image Quality Database: Eye-Tracking Release 1* [Online]. Available: http://mmi.tudelft.nl/iqlab/eye_tracking_1.html

[2] Z. Wang and A. C. Bovik, "Modern image quality assessment," in *Proc. Synthesis Lectures Image, Video, Multimedia Process.*, 2006, p. 156.

[3] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.

[4] S. Daly, "The visible difference predictor: An algorithm for the assessment of image fidelity," in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp. 179–206.

[5] J. Lubin, "The use of psychophysical data and models in the analysis of display system performance," in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp. 163–178.

[6] R. J. Safranek and J. D. Johnston, "A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression," in *Proc. IEEE Int. Conf. Acou., Speech Signal Process.*, May 1989, pp. 1945–1948.

[7] A. B. Watson, J. Hu, and J. F. Mcgowan, "DVQ: A digital video quality metric based on human vision," *J. Electron. Imag.*, vol. 10, no. 1, pp. 20–29, 2001.

[8] H. R. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," *IEEE Signal Process. Lett.*, vol. 4, no. 11, pp. 317–320, Nov. 1997.

[9] Z. Wang, A. C. Bovik, and B. L. Evans, "Blind measurement of blocking artifacts in images," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3. Sep. 2000, pp. 981–984.

[10] H. Liu and I. Heynderickx, "A perceptually relevant no-reference blockiness metric based on local image characteristics," *EURASIP J. Adv. Signal Process.*, vol. 2009, no. 263540, pp. 1–14, 2009.

[11] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3. Sep. 2002, pp. 57–60.

[12] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 717–728, Apr. 2009.

[13] H. Liu, N. Klomp, and I. Heynderickx, "A no-reference metric for perceived ringing artifacts in images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 4, pp. 529–539, Apr. 2010.

[14] 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.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[15] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[16] H. R. Sheikh, A. C. Bovik, and L. K. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1918–1927, Dec. 2005.

[17] R. V. Babu, S. Suresh, and A. Perkis, "No-reference JPEG-image quality assessment using GAP-RBF," *Signal Process.*, vol. 87, no. 6, pp. 1493–1503, 2007.

[18] P. Gastaldo and R. Zunino, "Neural networks for the no-reference assessment of perceived quality," *J. Electron. Imag.*, vol. 14, no. 3, pp. 1–11, 2005.

[19] R. Barland and A. Saadane, "Blind quality metric using a perceptual importance map for JPEG-2000 compressed images," in *Proc. IEEE Int. Conf. ICIP*, Oct. 2006, pp. 2941–2944.

[20] D. V. Rao, N. Sudhakar, I. R. Babu, and L. P. Reddy, "Image quality assessment complemented with visual region of interest," in *Proc. Int. Conf. Comput.: Theory Applicat.*, 2007, pp. 681–687.

[21] Q. Ma and L. Zhang, "Image quality assessment with visual attention," in *Proc. ICPR*, Dec. 2008, pp. 1–4.

[22] N. G. Sadaka, L. J. Karam, R. Ferzli, and G. P. Abousleman, "A no-reference perceptual image sharpness metric based on saliency-weighted foveal pooling," in *Proc. IEEE Int. Conf. ICIP*, Oct. 2008, pp. 369–372.

[23] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE J. Select. Topics Signal Process.* (Special Issue on Visual Media Quality Assessment), vol. 3, no. 2, pp. 193–201, Apr. 2009.

[24] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[25] U. Rajashekar, A. C. Bovik, and L. K. Cormack, "Gaffe: A gaze-attentive fixation finding engine," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 564–573, Apr. 2008.

[26] A. L. Yarbus, *Eye Movements and Vision*. New York: Plenum Press, 1967.

[27] A. Ninassi, O. L. Meur, P. L. Callet, and D. Barba, "Does where you gaze on an image affect your perception of quality? Applying visual attention to image quality metric," in *Proc. IEEE Int. Conf. ICIP*, Oct. 2007, pp. 169–172.

[28] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document ITU-R Rec. BT.500-11, International Telecommunication Union, Geneva, Switzerland, 2002.

[29] C. T. Vu, E. C. Larson, and D. M. Chandler, "Visual fixation patterns when judging image quality: Effects of distortion type, amount, and subject experience," in *Proc. IEEE SSIAI*, Mar. 2008, pp. 73–76.

[30] E. C. Larson, C. T. Vu, and D. M. Chandler, "Can visual fixation patterns improve image fidelity assessment?" in *Proc. Int. Conf. Image Process.*, Oct. 2008, pp. 2572–2575.

[31] U. Engelke and H.-J. Zepernick, "Framework for optimal region of interest-based quality assessment in wireless imaging," *J. Electron. Imag.*, vol. 19, no. 1, article 011005, 2010.

[32] N. Ouerhani, R. V. Wartburg, H. Hugli, and R. Muri, "Empirical validation of the saliency-based model of visual attention," *Electron. Lett. Comput. Vision Image Anal.*, vol. 3, no. 1, pp. 13–24, 2004.

[33] D. D. Salvucci, "A model of eye movements and visual attention," in *Proc. 3rd Int. Conf. Cognit. Model.*, 2000, pp. 252–259.

[34] C. Privitera and L. Stark, "Algorithms for defining visual regions-of-interest: Comparison with eye fixations," *Patt. Anal. Mach. Intell.*, vol. 22, no. 9, pp. 970–981, 2000.

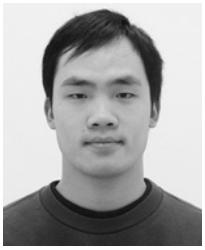
[35] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. *LIVE Image Quality Assessment Database Release 2* [Online]. Available: http://live.ece.utexas.edu/research/quality

[36] New Delft Experience Lab [Online]. Available: http://mmi.tudelft.nl/experiencelab

[37] H. Alers, J. Redi, H. Liu, and I. Heynderickx, "Task effects on saliency: Comparing attention behavior for free-looking and quality-assessment tasks," to be published.

[38] Video Quality Experts Group. *Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment Phase II* [Online]. Available: http://www.vqeg.org

[39] S. Winkler, "Vision models and quality metrics for image processing applications," Ph.D. dissertation, Dept. Elect., École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2002.
[40] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.
[41] D. C. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers*. New York: Wiley-Interscience, 1999.
[42] J. Redi, H. Liu, P. Gastaldo, R. Zunino, and I. Heynderickx, "How to apply spatial saliency into objective metrics for JPEG compressed images?" in *Proc. IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 961–964.
[43] U. Engelke, H. Liu, H.-J. Zepernick, I. Heynderickx, and A. Maeder, "Comparing two eye-tracking databases: The effect of experimental setup and image presentation time on the creation of saliency maps," in *Proc. Picture Coding Sym.*, 2010, pp. 282–285.



Hantao Liu (S'07–M'11) was born in Chengdu, China, in 1981. He received the M.S. degree in signal processing and communications from the University of Edinburgh, Edinburgh, Scotland, in 2005. Since 2006, he has been a Ph.D. student with the Department of Mediamatics, Delft University of Technology, Delft, The Netherlands.

He is currently working on a research project supported by the Philips Research Laboratories, Eindhoven, The Netherlands, developing no-reference objective metrics for perceived artifacts in compressed images. His current research interests include image analysis, visual perception, and signal processing.

Ingrid Heynderickx received the Ph.D. degree in physics from the University of Antwerp, Antwerp, Belgium, in December 1986.

In 1987, she joined the Philips Research Laboratories, Eindhoven, The Netherlands, and meanwhile worked in different areas of research such as optical design of displays, processing of liquid crystalline polymers, and functionality of personal care devices. Since 1999, she has been the Head of the research activities on visual perception of display and lighting systems. In 2005, she was appointed Research Fellow in the Visual Experiences Group. She is a member of the Society for Information Displays (SID), and for the SID, she was the Chairman of the Applied Vision Subcommittee from 2002 to 2007. In 2008, she became the Fellow of the SID and the Chairman of the European Committee of the SID. In 2005, she was appointed Guest Research Professor with the Southeast University of Nanjing, Nanjing, China, and part-time Full Professor with the University of Technology, Delft, The Netherlands.