Ulrich Engelke, Hagen Kaprykowsky, Hans-Jürgen Zepernick, and Patrick Ndjiki-Nya

Visual Attention in Quality Assessment

Theory, advances, and challenges

© INGRAM PUBLISHING

Perceptual quality metrics are widely deployed in image and video processing systems. These metrics aim to emulate the integral mechanisms of the human visual system (HVS) to correlate well with visual perception of quality. One integral property of the HVS is, however, often neglected: visual attention (VA) [1]. The essential mechanisms associated with VA consist mainly of higher cognitive processing, deployed to reduce the complexity of scene analysis. For this purpose, a subset of the visual information is selected by shifting the focus of attention across the visual scene to the most relevant objects. By neglecting VA, perceptual quality models inherently assume that all objects draw the attention of the viewer to the same degree. This applies to both the natural scene content as well as possibly induced distortions. However, suprathreshold distortions can be a strong attractor of VA and as a result, have a severe impact on the perceived quality. Identifying the perceptual influence of distortions relative to the natural content can thus be expected to enhance the prediction performance of perceptual quality metrics. The potential benefit of integrating VA information into image and video quality models has recently been recognized by a number of research groups [2]–[20]. The conclusions drawn from these works are somewhat controversial and give rise to many open questions. The goals of this article are therefore to shed some light onto this immature research field and to provide guidance for further advances. Toward these goals, we first discuss VA concepts that are relevant in the context of quality perception. We then review recent advances in research on integrating VA into quality assessment and

Digital Object Identifier 10.1109/MSP.2011.942473
Date of publication: 1 November 2011

1053-5888/11/$26.00©2011IEEE

highlight the main findings. Finally, we discuss major challenges and suggest potential solutions and future directions.

VISUAL ATTENTION

The human eye faces an abundant amount of visual information at any instant in time. Several mechanisms in early vision and higher cognitive layers are therefore deployed to reduce the complexity of scene analysis.

RETINAL SAMPLING AND EYE MOVEMENTS

Nonuniform sampling is deployed on the retina with a high sampling density in the fovea and rapidly diminishing density with increasing eccentricity. Hence, high-accuracy processing is limited to the central focus point, the fovea, and the peripheral visual field is perceived with lower accuracy.

A visual scene is gradually inspected by shifting the focus point using rapid, saccadic eye movements to fixate on the most relevant information in any context. Visual perception is active only during fixations and is largely suppressed during saccades [21]. Even though visual scene sampling is dominated by fixations, the scene is perceived as a continuous visual world. Fixations on moving objects are enabled through smooth pursuit eye movements, during which high acuity processing is performed for object speeds of up to approximately two degrees of visual angle per second [22].

VISUAL ATTENTION MECHANISMS

VA is thought to have evolved as a result of the limited overall resources available in the HVS [23]. Visual stimuli are therefore constantly competing for these resources and the most relevant stimuli in a given context are favored over the less relevant ones. The quality and acuity of the attended stimulus is enhanced through increased gain and contrast sensitivity, accompanied by a widespread baseline-activity reduction and noise suppression in the remaining visual field [24]. The decision which stimuli are favored is influenced by a number of different mechanisms.

OVERT VERSUS COVERT ATTENTION

Eye movements do not necessarily reflect exactly what human observers are attending to [25]. Overt VA relates to the act of directing the eyes to a stimulus whereas covert VA is related to a mental shift of attention. Covert attention precedes eye movements [26] and during fixation, it can be deployed to multiple locations simultaneously. Hence, covert VA allows us to efficiently monitor the visual scene and guide our eye movements. It is even possible to pursue one target while attending another target with only little effect on the pursuit [27]. Overt and covert VA are strongly interlinked and thus, eye tracking experiments are widely used to measure overt VA of human observers to gain insights into the attentive behavior.

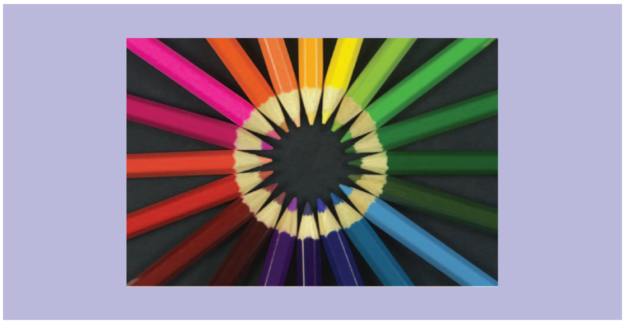
SPATIAL, FEATURE-BASED, AND OBJECT-BASED ATTENTION

VA is strongly influenced by three cues that are deployed simultaneously in a mutually optimal way; spatial location, low-level features, and objects [23]. Overt spatial attention is accompanied by eye movements whereas covert spatial attention can be deployed in the peripheral visual field and is thus not directly observable. Feature-based attention is largely independent of location and is affected by low-level features that are visually salient, including color, motion, orientation, and size [25]. It is active simultaneously throughout the visual field and is thus instrumental in improving detection performance of relevant stimuli. Object-based attention is guided by higher-level features, such as object structures as well as semantic information and contextual effects. Context plays a particularly important role in the decision process as to which object is considered more relevant than others [26].

BOTTOM-UP AND TOP-DOWN MECHANISMS

VA is guided by two main mechanisms: bottom-up and top-down. The former is reflexive, signal driven, and independent of a particular task. Bottom-up attention is fast, short lasting (transient), and performed in a preattentive manner across the visual field. It is driven involuntarily as a response to certain low-level features that are experienced as visually salient and distinct from the background. Motion, and in particular sudden temporal changes, are known to be dominant features in dynamic visual scenes [28], [29]. Motion increases the processing cost of visual perception and as a result of limited processing power in the HVS, considerably reduces visual sensitivity. This phenomenon, referred to as motion suppression [30], happens mainly in low-attentional areas when motion is different to that in high-attentional areas.

Top-down attention, on the other hand, is driven by higher-level cognitive factors and external influences, such as, semantic information, contextual effects, viewing task, and personal preference, expectations, experience and emotions. Top-down attention is slower, longer lasting (sustained), and unlike bottom-up attention, it requires a voluntary effort to shift the gaze. Top-down attention is considered to have a modulatory effect on bottom-up attention [31]. This is illustrated with regard to Figure 1. When shown this image, the attention of different observers would be driven to different pencils (bottom-up). However, if given the search task to identify the light blue pencil, the attention would be drawn to the pencil in the bottom



[FIG1] Illustration of the modulatory effect of top-down attention on bottom-up attention (image "coloring pencils" courtesy of [32]).

right corner. It is, however, extremely difficult for observers to ignore transient cues and hence, bottom-up attention is highly dominant in situations where there is a sudden onset of a visual stimulus. This phenomenon occurs independent of the task and is referred to as attentional capture.

BOTTOM-UP MODELS ONLY PERFORM WELL ON VISUAL SCENES THAT DO NOT CONTAIN ANY SEMANTIC INFORMATION OR ANY INTERESTING AND MEANINGFUL OBJECTS, WHICH IS RARELY THE CASE IN NATURAL IMAGE AND VIDEO CONTENT.

COMPUTATIONAL VISUAL ATTENTION MODELING

Computational VA models aim to predict the gaze locations of human observers. Current models are inspired by early works such as the feature integration theory by Treisman and Gelade [33], guided search by Wolfe et al. [34], or neural-based architecture by Koch and Ullman [35]. The latter model especially constituted a theoretical basis for biologically plausible models incorporating low-level characteristics of the HVS known to contribute to VA, such as multiple-scale processing, contrast sensitivity, and center-surround processing. A recent trend in computational saliency modeling is the development of statistical [36], information theoretic [37], and Bayesian approaches [26], [28], [38]. A main strength of these models is a strong mathematical foundation.

BOTTOM-UP MODELING

The majority of models focus on bottom-up mechanisms to predict visually salient locations [37]–[42]. Common traits of these models are a feature extraction stage followed by a not yet well-understood pooling into a final conspicuity map. Different feature combination strategies were investigated in [43]. The best tradeoff between prediction performance and generalization was achieved by nonlinear competition between salient locations followed by summation. An interesting additive feature integration method is proposed in [44]. In addition to the contribution of individual features, coupling factors were derived from psychophysical evidence to account for complex interactions between features.

A recent study [45] compared the saliency prediction performance of 13 bottom-up models. It was found that the maximum rather than the average predicted saliency correlates considerably better with human saliency recordings. Two models based on multiple-scale contrast-based processing were found to perform best in predicting visual saliency. Despite these findings, it is to date not fully understood how different feature dimensions contribute to overall visual saliency. None of the 13 tested models, for instance, accounts for momentary eye fixations and thus, variations in visual resolution on the retina. Taking into account whether a target can be identified from distractors in peripheral vision (known as crowding effect [46]) is of great concern in visual search tasks though.

Peripheral vision is highly sensitive to temporal activities [47], thus enhancing detection and perception of temporal changes across the visual field. Motion is therefore among the most dominant features to attract attention and thus needs to be an integral feature of any VA model in the context of dynamic visual scenes [29], [36], [48]–[50]. The models in [48] and [49] compute spatial and temporal features independently and fuse them in a pooling stage. Assuming that spatial and motion cues are not separable, the nonparametric models in [50] and [36] outperform earlier models by computing spatiotemporal features based on the phase-spectrum and spatiotemporal local steering kernels, respectively. A biologically inspired spatiotemporal saliency model based on a center-surround framework is proposed in [29]. The incorporation of spatiotemporal aspects into VA models is still an open issue, primarily since human perception of dynamic scenes lacks a theoretical foundation, as is available for still images. From a computational modeling viewpoint, one major challenge is to account for the various combinations of static to dynamic egomotion and scene motion in natural video sequences.

Bottom-up models only perform well on visual scenes that do not contain any semantic information or any interesting and meaningful objects, which is rarely the case in natural image and video content. Furthermore, bottom-up models process the visual scene in a local-to-global manner, meaning, that local features are accumulated into global conspicuity maps. According to this strategy, the number of candidate targets can be high and the scanpath prediction is difficult. A more recent holistic approach shows that the gist of a visual scene is perceived preattentively and can therefore already be integrated prior to the first saccade [26].

TOP-DOWN MODELING

Bottom-up and top-down cues need to be fused in a meaningful way to obtain a single focus of attention. Several works have tackled the difficult task of integrating top-down information with bottom-up features [26], [51]–[53]. A Bayesian framework for contextual guidance is proposed in [26], which is based on parallel computation of local saliency and global context features that enhance object and scene change detection. In visual search tasks, prior knowledge about the target is of particular importance as it strongly influences the search performance (see the coloring pencils example in the section "Bottom-Up and Top-Down Mechanisms"). Therefore, the target-relevant region should be excited, the target-irrelevant regions inhibited, or a combination thereof [54]. The performance of the well-known bottom-up model by Itti et al. [39] was improved by taking into account top-down cues to enable visual search. The degree to which these mechanisms contribute to the overall model needs to be adaptive to the current situation, with bottom-up cues dominating in exploratory (free viewing) conditions and top-down cues dominating in visual search tasks.

Independent of the viewing conditions, neither of the two mechanisms should be entirely suppressed.

IN VISUAL SEARCH TASKS, PRIOR KNOWLEDGE ABOUT THE TARGET IS OF PARTICULAR IMPORTANCE AS IT STRONGLY INFLUENCES THE SEARCH PERFORMANCE.

VISUAL ATTENTION FOR QUALITY ASSESSMENT: RECENT ADVANCES

Increased awareness to the strong interaction between VA and quality perception led to a number of computational methods that integrate VA into quality metrics to potentially improve prediction performance. We discuss in the following the most common VA integration methods and review recent advances for image [2]–[9] and video applications [10]–[20].

COMMON VISUAL ATTENTION INTEGRATION METHODS

We categorize the most common VA integration methods as illustrated in Figure 2. In Method 1, the perceptual difference (PD) between a test (T) and reference stimulus (R) is evaluated independently from the natural scene saliency. In a pooling stage, the perceptual difference is then typically weighted using the saliency map (SM), yielding the final quality score (Q). Assuming that distortions alter attention, the saliency difference (SD) between the reference and distorted stimuli can be used instead of or in addition to the natural scene saliency. Models following Method 2 first segment the image or video frames into salient regions (S) and background (B) using natural scene saliency. The perceptual difference is then computed independently on these regions and combined into an overall quality metric using a weighted summation.

IMAGE QUALITY ASSESSMENT

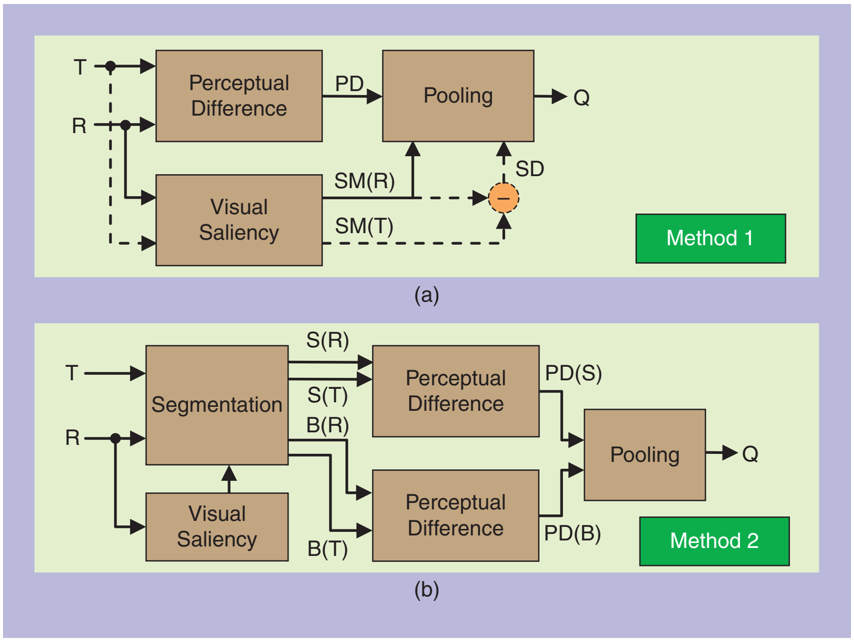
Method 1 is the most widely adopted approach in image quality assessment [2]–[7]. Despite the common integration method, different conclusions arise from these works.

Barland et al. [2] used the Osberger VA model [48] and proposed a multiple-scale VA model for integration into no-reference (NR) blur and ringing metrics for JPEG2000 compressed images. The proposed model yielded a superior performance, which supports the finding in [45] that multiple-scale processing is beneficial for VA models. Sadaka et al. [4] integrated bottom-up saliency [39] into their sharpness metric through multiplicative weighting with the distortion map. The linear correlation coefficient (CC) was enhanced from CC = 0.58 to CC = 0.69. The rather low performance of the original metric, however, provided a big margin for improvement. Moorthy et al. [5] incorporate bottom-up saliency [41] into the structural similarity (SSIM) index [55]. An improvement of CC of approximately 1–4% was achieved across different distortions covered in the test images. No results are reported to validate the statistical significance of the rather low improvements.

Gkioulekas et al. [6] adopt the surprise model in [28] for images and incorporate it into SSIM through weighted summation. The authors found that their surprise model improves SSIM considerably more than the bottom-up model in [39]. It was further found that maximum local saliency provides superior results than averaged local saliency, which is in line with findings in [45], [56]. The improvements to the original SSIM index of approximately 1% in CC are marginal though.

Instead of using a VA model, Ninassi et al. [3] integrated fixation density maps (FDM) from quality-task eye tracking into SSIM and the mean absolute distance (MAD) metric. On the contrary to the other works, no improvements were found in the context of JPEG and JPEG2000 distorted images.

A comprehensive study on incorporating task-free and quality-task eye tracking data into quality metrics (SSIM, visual information fidelity (VIF) [57] criterion, peak signal-to-noise ratio (PSNR), generalized block edge impairment metric (GBIM) [58]) has recently been published by Liu et al. [7]. Statistically significant improvements were found for all metrics, with a superior performance in the case of task-free eye tracking data. The improvement was shown to be larger for images with distinct salient locations, as compared to images that have widely spread saliency. It was further concluded that background distortions should not be neglected, in particular in



[FIG2] Common methods of integrating VA into quality metrics: (a) Method 1 and (b) Method 2.

visual scenes where distortion visibility in the background is considerably higher than in the salient region.

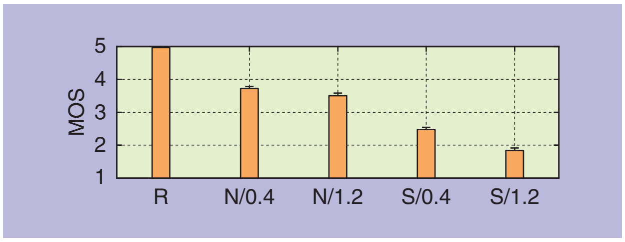
The suitability of Method 2 for VA integration was evaluated in [8], [9]. Larson et al. [8] segmented images into primary regions-of-interest (ROI), secondary ROI, and background based on task-free and quality-task eye tracking data. Five quality metrics [SSIM, VIF, PSNR, visual signal-to-noise ratio (VSNR) [59], and weighted signal-to-noise ratio (WSNR)] were computed independently on these regions and combined using a weighted summation. All metrics received highest weights for the primary ROI, apart from VSNR, which favored the secondary ROI. Superior improvement is reported with task-free rather than the quality-task eye tracking data. Unlike in [7], no improvements were found to be significant.

Engelke et al. [9] proposed an optimization framework for ROI-based image quality metrics in the context of wireless imaging distortions. Significant improvements were found for SSIM, VIF, and PSNR, which are believed to be due to the localized nature of the distortions. Unlike with global distortions that were considered in previous works [3], [5], [8], the impact of the distortion location inside or outside the ROI has a more significant impact. In line with [8], however, it was also found that VSNR received a higher weight for the background, which emphasizes the sensitivity of saliency integration to the distortion measure used.

VIDEO QUALITY ASSESSMENT

Due to the dynamic changes of the visual scene in video applications, it is usually impossible to observe all details within every frame. Our gaze is mainly driven to follow the most salient regions, unlike with images, where sufficiently long viewing times also allow to analyze the background regions. Distortions that occur outside the most salient areas are therefore assumed to have a lower impact on the overall quality and VA concepts can be expected to have a higher impact in video as compared to image applications. The strong attentional guidance due to motion cues and temporal changes plays an essential role and has been implemented in many VA integration methods, as discussed in the following.

Cavallaro et al. [10] integrated a low-level feature-based (motion, color) quality metric with extraction of semantic information by means of face segmentation. Independent quality assessment in the faces and the background, followed by a pooling stage, led to considerable improvements.



[FIG3] MOS for five different distortion classes: (R = reference, S = salient region, N = nonsalient region, 0.4 = 0.4 s distortion length, and 1.2 = 1.2 s distortion length).

Lu et al. [11] modulate the distortion maps of just-noticeable-difference (JND) models as well as PSNR and SSIM using an original VA model based on bottom-up (color, texture, motion) and top-down (faces, skin color) cues. The model accounts for absolute and relative motion as well as motion suppression. All features are pooled using the model in [44]. The JND and quality models were strongly improved.

You et al. [12] integrate top-down cues (faces and text) with the bottom-up model in [39]. Different weighting schemes are tested for VA integration into SSIM and PSNR. Improvements were found only for PSNR but not for SSIM, and it is concluded that SSIM is unsuitable for VA integration. Considering the finding in [7] it may be that this conclusion arises from unsuitable pooling that neglects distortions in the background of the visual scene. In [16], the same group takes into account global quality and motion in addition to local, saliency-based quality analysis. This combination is found to outperform the individual local and global quality measures.

Ma et al. [13] propose a complex VA model to weight spatial distortions without totally neglecting background distortions. In addition, motion suppression is accounted for as well as egomotion of the camera. Integration of the model into SSIM, VIF, and PSNR improved performance of the metrics approximately 8%, 5%, and 3%, respectively.

Engelke et al. [15] conducted a quality-task eye tracking experiment to identify the perceived annoyance of packet loss distortions located either in a salient (S) or nonsalient (N) region. Two different distortion lengths were considered as well (0.4 s and 1.2 s). The mean opinion scores (MOS) presented in Figure 3 reveal that distortions located in salient regions are considerably more annoying than distortions in nonsalient regions. In fact, even the short distortions in the salient regions (S/0.4) received one MOS unit lower than the long distortions in the nonsalient region (N/1.2). Based on these results, a saliency awareness framework for VQM in the context of localized packet loss distortions was proposed [14]. The contemporary temporal trajectory aware VQM (TetraVQM) [60] and PSNR could be improved by penalizing the distortion measures in relation to the underlying content saliency.

Le Meur et al. [17] integrated task-free and quality-task eye tracking data into an original VQM. No improvements were reported with either of the eye tracking data in the context of H.264/AVC compression distortions, which agrees with an earlier study of the same group on images [3]. However, as in the previous study, only simple spatial pooling functions have been considered for VA integration. In a similar study in the context of H.264 compression distortions, Gao et al. [18] report a 4% improvement in CC by integrating a spatiotemporal, bottom-up VA model into SSIM. However, no statistical significance analysis is provided to support the validity of the results. Generally, the global compression distortions considered in these studies can be assumed to have little effect on the VA integration in comparison to, for instance, the localized packet loss distortions considered in [14].

The study by Feng et al. [19] supports the strong impact of localized packet loss distortions in relation to content saliency. Unlike the previously discussed works, this study analyzed the potential benefits of taking into account the saliency difference (SD) between the reference and distorted video (see Method 1 in Figure 2). The intensity, color, and orientation features from the bottom-up model in [39] were extended with a motion model and subject to a weighted summation. By incorporating this model into SSIM, MAD, and the mean squared error (MSE), correlations with subjective quality ratings of up to 0.99 were achieved. Given the relatively large number of seven parameters in the model as compared to the training set of 12 sequences, the model may in fact be overfitted to some degree.

Ćulibrk et al. [20] took a different approach to the two methods depicted in Figure 2. Instead of combining an existing quality metric with a VA model, 35 different features were considered to train a regression tree. Only the five features that had the most significant impact on the metrics performance in the context of MPEG-2 compression distortions were selected. It was found that blocking and blur artifacts were highly annoying in the salient regions whereas temporal distortions were annoying throughout the visual field. The authors concluded that background distortions should not be entirely neglected for successful saliency integration into quality assessment, which supports the conclusions drawn in [7] for images.

SUMMARY OF FINDINGS

From the works discussed in this section, we can summarize that improvement in quality prediction performance due to VA integration is generally superior

1) in video rather than image applications
2) in the case of localized rather than global distortions
3) with task-free instead of quality-task eye tracking data
4) if top-down cues are integrated in addition to bottom-up cues (thus far, usually only faces and text are considered)
5) if motion (relative motion, motion suppression, egomotion) is appropriately integrated in video applications
6) if background distortions are not entirely suppressed but only relative to salient region distortions
7) if multiple-scale analysis is included in the VA model.

Despite many common agreements, some conclusions about the potential benefits of VA integration for quality assessment are controversial. For instance, in [12] it was concluded that SSIM is unsuitable for saliency inclusion whereas [5], [7], and [9] reported particularly good improvements for SSIM. Such controversies are an indication of the many challenges that still need to be solved. Some of the major challenges are identified and discussed in the following section.

CURRENT CHALLENGES

GROUND TRUTH SELECTION

The kind of VA ground truth incorporated into the quality metrics is assumed to have a strong impact on the success of the VA integration performance. Several aspects are of particular interest in this respect.

COMPUTATIONAL MODELS VERSUS PSYCHOPHYSICAL DATA

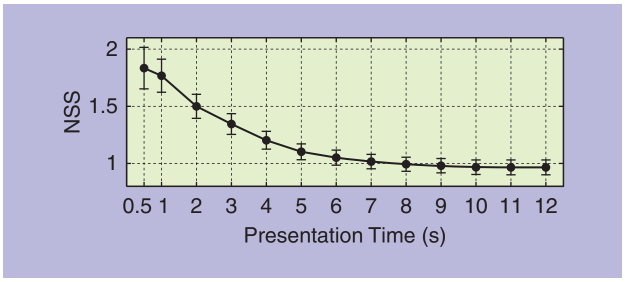
The saliency ground truth used throughout the works discussed in the section "Visual Attention for Quality Assessment: Recent Advances" is either based on computational VA models or psychophysical data. The former has the advantage that it enables automated deployment in image and video processing systems. However, even current state-of-the-art VA models are not reliable predictors of human viewing behavior, often because they focus on bottom-up cues, neglecting important top-down cues and semantic information. Psychophysical experiments, on the other hand, are considered to be a reliable ground truth. To avoid potential modeling errors due to poorly performing VA predictors, we therefore strongly recommend to use psychophysical data as ground truth. As psychophysical experiments find no deployment in real-time applications, it is of great concern to develop VA models that more reliably predict human viewing behavior.

EYE MOVEMENTS VERSUS REGIONS-OF-INTEREST

Eye tracking data is the most common psychophysical ground truth. As eye movements are driven by bottom-up and top-down cues, it is difficult to identify to which degree low-level features and object-level semantics contribute to the resulting FDM. Furthermore, during the search for the most interesting or informative regions, humans do not only attend useful locations [22] and as such, eye tracking recordings do not provide direct insight into which regions are of interest. To obtain more direct insight into perceived interest and its interrelation with eye movements, we conducted an experiment in which human observers hand-labeled ROI in natural images [61]. The resulting ROI maps are compared to FDM from an eye tracking experiment [62] with the aim to identify the presentation time that best predicts the ROI maps. The degree of similarity between FDM and ROI selections is quantified using the normalized scanpath saliency (NSS) [63], as presented in Figure 4. The similarity gradually decreases with presentation time during eye tracking, which suggests that early fixations best predict the ROI. Similar results were reported in [64] and [65]. These findings support the earlier discussion that the gist of a scene is perceived preattentively and thus guides early eye movements (see the section "Bottom-Up Modeling"). The conclusions are expected to be highly task dependent though, since, for instance, a radiologist searching for breast cancer would unlikely attend the target with the early fixations. Similar studies are needed for video, as bottom-up motion cues strongly guide attention and thus might lead to different conclusions.

TASK-FREE VERSUS QUALITY-TASK EYE TRACKING DATA

Whether to use eye tracking data from task-free or quality assessment condition as a ground truth is still an open question. Viewing behavior can change considerably in a visual search task such as quality assessment [51], [66].



[FIG4] NSS between FDM and ROI selections for different presentation times.

Covert VA improves speed and accuracy on many detection, discrimination and localization tasks [23], for which reason observers are sensitized to distortions during quality assessment. This is particularly true since the observer usually has prior knowledge about the distortions (the target). Models that aim to predict gaze patterns recorded under quality assessment task therefore need to be tuned accordingly and attention to distortions needs to be excited relative to the content.

In natural conditions, humans do not view images or video sequences with the aim to identify possible degradations in the content. Their attention is therefore not sensitized to these targets. As the ultimate goal of quality assessment is the prediction of quality perception during these natural conditions, eye tracking data from task-free experiments might in fact be the more sensible choice. The validity of these presumptions is believed to hold particularly for static visual scenes, for images, and is supported by several recent studies [3], [7], [8]. It was found that fixations spread more into the background of the visual scene and thus overestimate the relative impact of distortions in the background to distortions in salient regions [7]. In dynamic visual scenes, on the other hand, fixation durations and locations were not found to be significantly different between task-free and quality assessment conditions [17]. This can be largely explained through the phenomenon of attentional capture due to motion and temporal changes, as discussed in the section "Bottom-Up and Top-Down Mechanisms." In summary, the choice between a task-free and quality assessment task is potentially more crucial in image as compared to video quality assessment. More studies are needed to confirm these observations.

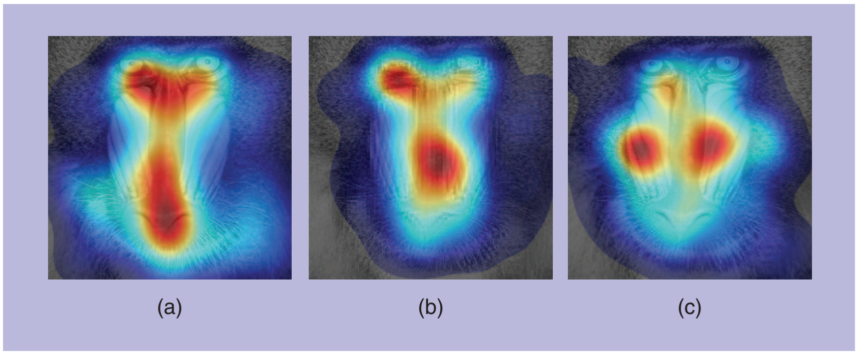
IMPACT OF DISTORTIONS ON VISUAL ATTENTION

The degree to which distortions attract attention in relation to the underlying natural image or video content depends on many influencing factors, such as the natural content saliency and the distortion type, strength, and distribution. In general, distortions that are strongly salient compared to the natural content are expected to attract more attention and thus result in a stronger impact on the overall perceived quality.

GLOBAL VERSUS LOCAL DISTORTIONS

Spatially and spatiotemporally local distortions (e.g., due to packet loss) were shown to attract attention comparably stronger than globally distributed distortions (e.g., due to compression) [19], [67], [68]. This is related to the Bayesian notion of surprise [28], which states that novel events resemble saliency in space and in time and are thus strong attention attractors. High temporal sensitivity in peripheral vision further supports detection of local and time varying distortions. Local distortions therefore alternate gaze patterns relatively strong compared to global distortions. Recent psychophysical evidence supports this rationale. In [69], it was found that global compression distortions do not alter viewing patterns considerably while in [68] it was shown that localized packet loss distortions considerably change viewing behavior.

We studied the shift of gaze patterns during image quality assessment in the case of localized wireless imaging distortions. Figure 5(a) depicts a heat map on the undistorted "Mandrill" image. The distorted versions in Figure 5(b) and (c) exhibit strong blocking distortions and subtle ringing distortions, respectively. Against intuition, the subtle ringing distortions change the gaze pattern considerably more than the strong blocking distortions. Despite the stronger shift, the image in Figure 5(c) received a considerably higher MOS of 64 (on a scale from zero to 100) compared to 25 for the image in Figure 5(b). Covert attention shifts between reference and distorted images thus need to be handled with great caution, as they do not directly relate to quality perception. These observations confirm the earlier discussion that quality-task eye tracking data is unsuitable in case of images, in particular in case of localized distortions.



[FIG5] Heat maps for the image "Mandrill": (a) reference image, (b) image with strong blocking artifacts, and (c) image with subtle ringing artifacts.

DISTORTIONS IN RELATION TO CONTENT SALIENCY

The alternation of viewing behavior was found to be strongly depending on whether distortions are appearing in salient or nonsalient regions [68]. This phenomenon is illustrated for video in Figure 6. The area under the receiver operating characteristic (ROC) curve (AUC) is used to measure the amount of overt attention in the respective distortion regions (salient or nonsalient). As expected, the nonsalient regions are attended less than the salient regions throughout the video sequence. Upon appearance of the packet loss distortions (A), the gaze is shifted towards the distortions in the nonsalient region, as indicated by the rise in AUC in Figure 6(a). After disappearance (D), the gaze shifts back to the salient region. Unlike in the case of images, we found that the MOS were highly correlated with the AUC in the distortion regions (CC $= -0.79$). Thus, attention to distortions in video is indeed related to the overall perceived quality, even under quality-task condition.

The attention shift, however, was not observed for all sequences, as indicated in Figure 6(b). The attention shift appears to be strongly dependent on the relative strength of saliency between content and distortions. The relative location is also important as performance in visual search tasks deteriorates with increased eccentricity in peripheral vision [23]. These findings provide only indications of the complex interaction between content and distortion saliency. Task-free eye tracking experiments on distorted content are needed to study distortion related attention shifts under natural viewing conditions.

VISUAL ATTENTION INTEGRATION

The pooling of the VA and distortion information is probably the most crucial step of VA integration into quality metrics. The psychophysiological mechanisms underlying the interaction between VA and quality perception are not well understood yet. Given the recent psychophysical findings, however, some interesting directions for improved pooling methods can be derived.

SOME GENERAL ISSUES

The combination of model parameters is often done in an ad hoc manner, using simple, Minkowski-like pooling functions. The pooling step typically introduces additional parameters to the model and thus, allows for the designer to better fit the model to the data. A theoretical foundation about pooling methods is needed to comprehend to what degree the improvement is due to saliency integration or due to increased degrees of freedom alone.

Many studies performed are still using purely bottom-up VA models, even though they are known to not perform well in complex natural scenes. Top-down models therefore need to be included into the pooling stage. Additive rather than multiplicative pooling should be used [54], since both bottom-up and top-down cues influence viewing behavior in any context and should therefore not be suppressed entirely.

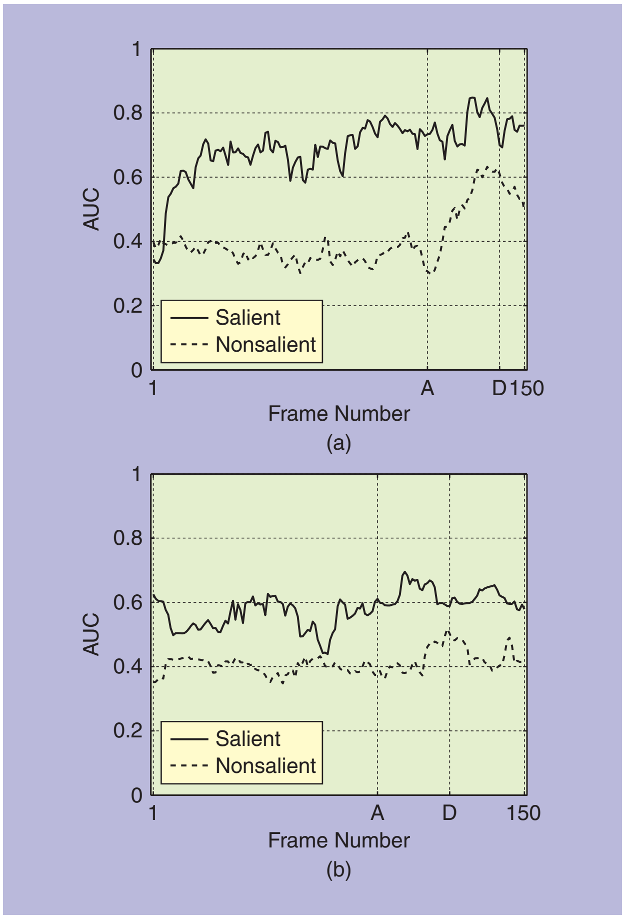
SPATIAL POOLING

Image distortions have perceptual impact whether they are in the salient region or not, especially in the context of local distortions. Typical pooling steps, which multiply saliency maps with distortion maps, often suppress background distortions entirely. Depending on the masking properties of the image, background distortions can be strong attractors of attention and are perceived as highly annoying [7]. The pooling method in [13] accounts for background distortions and might constitute a good basis to exploit appropriate pooling of salient region and background distortions.

[BOTTOM-UP AND TOP-DOWN CUES INFLUENCE VIEWING BEHAVIOR IN ANY CONTEXT AND SHOULD THEREFORE NOT BE SUPPRESSED ENTIRELY.]

SPATIOTEMPORAL POOLING

Motion and temporal changes in video have a substantial impact on distortion perception. Due to motion suppression, detection and perception of distortions are considerably reduced in peripheral vision. Spatiotemporal contrast sensitivity functions used in video quality models should therefore be adapted in relation to the motion observed in the visual scene. Attentional capture, on the other hand, counteracts this phenomenon, causing easy detection of spatiotemporally local distortions in the background. Spatiotemporal distortions in the peripheral visual field should therefore not be entirely neglected. Taken these phenomena into account conjointly constitutes a great challenge and requires more sophisticated spatiotemporal VA models.



[FIG6] AUC of two video sequences for (a) a strong and (b) a weak attention shift towards the distortions.

TOWARD MORE APPROPRIATE POOLING METHODS

In most works reviewed in the section “Visual Attention for Quality Assessment: Recent Advances,” perceptual distortions and visual saliency are evaluated independently and combined in a pooling stage (see Figure 2). The strong interaction between VA to natural content and distortions, however, might call for more integrated methods that take into account content and distortion saliency simultaneously. In addition, other image and video properties, such as masking effects, need to be accounted for conjointly. To fully understand these interactions and to develop them into advanced pooling techniques, theoretical foundations need to be established first and more psychophysical evidence is needed. Some advances on bottom-up feature integration have been reported in the vision science community [43], [44]. The model in [44] takes into account feature interactions and might thus constitute a suitable candidate for the pooling stage. However, these experiments were carried out on simple stimuli and similar studies are needed for static and dynamic natural scene content.

CONCLUSIONS AND FUTURE DIRECTIONS

The current state of research discussed in this article suggests that there is indeed a benefit of integrating VA into perceptual quality assessment. Most notably, VQM for the assessment of localized artifacts may benefit from the incorporation of VA. However, the existing methods are strongly engineering inspired and the interaction between VA and quality perception is often simplified. Closer collaboration between the image processing and vision science communities is imperative to further enhance this immature field of research.

The following exciting issues were outside the scope of this article but are worth exploring. Most of the works discussed here were based on full-reference quality assessment. VA models are designed to work without any reference and may therefore provide valuable guidance to further develop no-reference quality metrics. Gaze patterns from eye tracking experiments are known to reflect predominantly overt VA. Psychophysiological data, such as through electroencephalography, needs to be investigated to obtain a better understanding of covert VA to distortions in natural content. In the context of multimedia, VA is driven not only by visual cues. Auditory cues are known to be a strong attractor of VA and their impact on attention deployment needs to be explored [70]. Upcoming three-dimensional (3-D) applications constitute an exciting research direction, since additional 3-D cues influence the attention of an observer. These applications induce their own range of distortions, each of them attracting attention to a certain degree that has yet to be investigated.

AUTHORS

Ulrich Engelke (ulrichengelke@gmail.com) received the Dipl.-Ing. degree in electrical engineering in 2004 from RWTH Aachen University, Germany, and the Ph.D. degree in telecommunications in 2010 from the Blekinge Institute of Technology, Sweden. His Ph.D. studies were largely funded by a five-year scholarship awarded through the Royal Institute of Technology (KTH), Sweden. In 2011, he pursued a postdoc position at the University of Nantes, France. Currently he is with the Visual Experiences Group at Philips Research, The Netherlands, working with perception in lighting applications. His research interests include visual scene understanding, human perception, psychophysical experimentation, and signal processing.

Hagen Kaprykowsky (hagen.kaprykowsky@gmail.com) received a German-French double-diploma degree in electrical engineering from the University of Karlsruhe (TH) and INP Grenoble in 2005. Within his diploma thesis, he developed a globally optimal dynamic time-warping algorithm for musical alignment at the IRCAM Centre Pompidou in Paris. He gained his first professional experiences at the German Research Center for Artificial Intelligence in Kaiserslautern, where he worked on the development of adaptive statistical methods for optical character recognition. In 2008 he joined the Image Processing Department of the Fraunhofer Heinrich Hertz Institute. His research interests include perceptual quality assessment and perception-oriented video coding.

Hans-Jürgen Zepernick (hans-jurgen.zepernick@bth.se) received the Dipl.-Ing. degree from the University of Siegen in 1987 and the Dr.-Ing. degree from the University of Hagen in 1994. From 1987 to 1989, he was with Siemens AG, Germany. He is currently a professor of radio communications at the Blekinge Institute of Technology, Sweden. His previous positions include professor of wireless communications at Curtin University of Technology; deputy director of the Australian Telecommunications Research Institute; and associate director of the Australian Telecommunications Cooperative Research Centre. His research interests include radio transmission techniques, mobile multimedia communications, and perceptual quality assessment.

Patrick Ndjiki-Nya (patrick.ndjiki-nya@hhi.fraunhofer.de) received the Dipl.-Ing. in 1997 and the Dr.-Ing. degree in 2008 from the Technische Universität Berlin. From 1997 to 1998, he was involved in the development of a flight simulator at Daimler-Benz. From 1998 to 2001 he was employed as development engineer at DSPecialists. During the same period, he researched content-based video features at Fraunhofer Heinrich Hertz Institute with the purpose of implementation in DSPecialists’ DSP solutions. Since 2001, he has been with Fraunhofer Heinrich Hertz Institute, where he was a project manager initially and senior project manager since 2004. He was appointed group manager in 2010. He is a Member of the IEEE.

REFERENCES

[1] J. Wolfe, “Visual attention,” in *Seeing,* K. K. D. Valois, Ed. San Diego, CA: Academic, 2000, pp. 335–386.

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

[3] 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,” in *Proc. IEEE Int. Conf. Image Processing*, Oct. 2007, vol. 2, pp. 169–172.

[4] 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. Image Processing*, Oct. 2008, pp. 369–372.

[5] A. K. Moorthy and A. C. Bovik, “Visual importance pooling for image quality assessment,” *IEEE J. Select. Topics Signal Processing*, vol. 3, no. 2, pp. 193–201, 2009.

[6] I. Gkioulekas, G. Evangelopoulos, and P. Maragos, “Spatial Bayesian surprise for image saliency and quality assessment,” in *Proc. IEEE Int. Conf. Image Processing*, Sept. 2010, pp. 1081–1084.

[7] H. Liu and I. Heynderickx, “Visual attention in objective image quality assessment: Based on eye-tracking data,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 7, pp. 971–982, July 2011.

[8] E. C. Larson, C. Vu, and D. M. Chandler, “Can visual fixation patterns improve image fidelity assessment?” in *Proc. IEEE Int. Conf. Image Processing*, Oct. 2008, pp. 2572–2575.

[9] U. Engelke and H.-J. Zepernick, “A framework for optimal region of interest-based quality assessment in wireless imaging,” *J. Electron. Imaging (Special Section on Image Quality)*, vol. 19, no. 1, 2010, pp. 1–13.

[10] A. Cavallaro and S. Winkler, “Segmentation-driven perceptual quality metrics,” in *Proc. IEEE Int. Conf. Image Processing*, Oct. 2004, vol. 5, pp. 3543–3546.

[11] Z. Lu, W. Lin, X. Yang, E. Ong, and S. Yao, “Modeling visual attention’s modulatory after effects on visual sensitivity and quality evaluation,” *IEEE Trans. Image Processing*, vol. 14, no. 11, pp. 1928–1942, 2005.

[12] J. You, A. Perkis, M. Hannuksela, and M. Gabbouj, “Perceptual quality assessment based on visual attention analysis,” in *Proc. ACM Int. Conf. Multimedia*, Oct. 2009, pp. 561–564.

[13] Q. Ma, L. Zhang, and B. Wang, “New strategy for image and video quality assessment,” *J. Electron. Imaging (Special Section on Image Quality)*, vol. 19, no. 1, 2010, pp. 1–14.

[14] U. Engelke, M. Barkowsky, P. Le Callet, and H.-J. Zepernick, “Modelling saliency awareness for objective video quality assessment,” in *Proc. Int. Workshop Quality Multimedia Experience*, June 2010, pp. 212–217.

[15] U. Engelke, R. Pepion, P. Le Callet, and H.-J. Zepernick, “Linking distortion perception and visual saliency in H.264/AVC coded video containing packet loss,” in *Proc. SPIE/IEEE Int. Conf. Visual Communications and Image Processing*, July 2010.

[16] J. You, J. Korhonen, and A. Perkis, “Attention modeling for video quality assessment: Balancing global quality and local quality,” in *Proc. IEEE Int. Conf. Multimedia and Expo*, July 2010, pp. 914–919.

[17] O. Le Meur, A. Ninassi, P. Le Callet, and D. Barba, “Overt visual attention for free-viewing and quality assessment tasks: Impact of the regions of interest on a video quality metric,” *Signal Process. Image Commun.*, vol. 25, no. 7, pp. 547–558, 2010.

[18] X. Gao, N. Liu, W. Lu, D. Tao, and X. Li, “Spatio-temporal salience based video quality assessment,” in *Proc. IEEE Int. Conf. Systems, Man and Cypernetics*, Oct. 2010, pp. 1501–1505.

[19] X. Feng, T. Liu, D. Yang, and Y. Wang, “Saliency inspired full reference quality metrics for packet-loss-impaired video,” *IEEE Trans. Broadcast.*, vol. 57, no. 1, pp. 81–88, Mar. 2011.

[20] D. Ćulibrk, M. Mirković, V. Zlokolica, M. Pokrić, V. Crnojević, and D. Kukolj, “Salient motion features for video quality assessment,” *IEEE Trans. Image Processing*, vol. 20, no. 4, pp. 948–958, Apr. 2011.

[21] D. Burr, M. C. Morrone, and J. Ross, “Selective suppression of the magnocellular visual pathway during saccadic eye movements,” *Nature*, vol. 371, no. 6497, pp. 511–513, 1994.

[22] E. Kowler, “Eye movements: The past 25 years,” *Vision Res.*, vol. 51, no. 13, pp. 1457–1483, 2011.

[23] M. Carrasco, “Visual attention: The past 25 years,” *Vision Res.*, vol. 51, no. 13, pp. 1484–1525, 2011.

[24] A. T. Smith, K. D. Singh, and M. W. Greenlee, “Attentional suppression of activity in the human visual cortex,” *Neuroreport*, vol. 11, no. 2, pp. 271–278, 2000.

[25] J. M. Wolfe and T. S. Horowitz, “What attributes guide the deployment of visual attention and how do they do it?” *Nature Rev. Neurosci.*, vol. 5, pp. 1–7, June 2004.

[26] A. Torralba, A. Oliva, M. Castelhano, and J. Henderson, “Contextual guidance of eye movements and attention in real-world scenes: The role of global features on object search,” *Psychol. Rev.*, vol. 113, no. 4, pp. 766–786, 2006.

[27] B. Khurana and E. Kowler, “Shared attentional control of smooth eye movement and perception,” *Vision Res.*, vol. 27, no. 9, pp. 1603–1618, 1987.

[28] L. Itti and P. Baldi, “Bayesian surprise attracts human attention,” *Vision Res.*, vol. 49, no. 10, pp. 1295–1306, 2009.

[29] V. Mahadevan and N. Vasconcelos, “Spatiotemporal saliency in dynamic scenes,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 32, no. 1, pp. 171–177, Jan. 2010.

[30] B. J. Murphy, “Pattern thresholds for moving and stationary gratings during smooth eye movement,” *Vision Res.*, vol. 18, no. 5, pp. 521–530, 1978.

[31] S. Treue, “Visual attention: The where, what, how and why of saliency,” *Curr. Opin. Neurobiol.*, vol. 13, no. 4, pp. 428–432, 2003.

[32] M. Maggs. (2007). Colouring pencils. [Online]. Available: http://commons.wikimedia.org/wiki/File:Colouring pencils.jpg

[33] A. M. Treisman and G. Gelade, “A feature-integration theory of attention,” *Cogn. Psychol.*, vol. 12, no. 1, pp. 97–136, 1980.

[34] 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.*, vol. 15, no. 3, pp. 419–433, 1989.

[35] C. Koch and S. Ullman, “Shifts in selection in visual attention: Towards the underlying neural circuitry,” *Hum. Neurobiol.*, vol. 4, no. 4, pp. 219–227, 1985.

[36] H. J. Seo and P. Milanfar, “Static and space-time visual saliency detection by self-resemblance,” *J. Vis.*, vol. 9, no. 12:15, pp. 1–27, 2009.

[37] N. D. B. Bruce and J. K. Tsotsos, “Saliency, attention, and visual search: An information theoretic approach,” *J. Vis.*, vol. 9, no. 3:5, pp. 1–24, 2009.

[38] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell, “SUN: A Bayesian framework for saliency using natural statistics,” *J. Vis.*, vol. 8, no. 7:32, pp. 1–20, 2008.

[39] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 11, pp. 1254–1259, 1998.

[40] 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. Machine Intell.*, vol. 28, no. 5, pp. 802–817, 2006.

[41] U. Rajashekar, I. van der Linde, A. C. Bovik, and L. K. Cormack, “GAFFE: A gaze-attentive fixation finding engine,” *IEEE Trans. Image Processing*, vol. 17, no. 4, pp. 564–573, 2008.

[42] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk, “Frequency-tuned salient region detection,” in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, June 2009, pp. 1597–1604.

[43] L. Itti and C. Koch, “Feature combination strategies for saliency-based visual attention systems,” *J. Electron. Imaging*, vol. 10, no. 1, pp. 161–169, 2001.

[44] H. C. Nothdurft, “Salience from feature contrast: Additivity across dimensions,” *Vision Res.*, vol. 40, no. 10–12, pp. 1183–1201, 2000.

[45] A. Toet, “Computational versus psychophysical bottom–up image saliency: A comparative evaluation study,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 99, 2011.

[46] D. M. Levi, “Crowding—An essential bottleneck for object recognition: A mini review,” *Vision Res.*, vol. 48, no. 5, pp. 635–654, 2008.

[47] S. P. McKee and K. Nakayama, “The detection of motion in the peripheral visual field,” *Vision Res.*, vol. 24, no. 1, pp. 25–32, 1984.

[48] W. Osberger and A. M. Rohaly, “Automatic detection of regions of interest in complex video sequences,” in *Proc. IS&T/SPIE Human Vision and Electronic Imaging VI*, Jan. 2001, vol. 4299, pp. 361–372.

[49] Y. Zhai and M. Shah, “Visual attention detection in video sequences using spatiotemporal cues,” in *Proc. ACM Int. Conf. Multimedia*, Oct. 2006, pp. 815–824.

[50] C. Guo, Q. Ma, and L. Zhang, “Spatio-temporal saliency detection using phase spectrum of quaternion Fourier transform,” in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, June 2008, pp. 1–8.

[51] F. Navalpakkam and L. Itti, “Modeling the influence of task on attention,” *Vision Res.*, vol. 45, no. 2, pp. 205–231, 2005.

[52] Y. F. Ma, X. S. Hua, L. Lu, and H. J. Zhang, “A generic framework of user attention model and its application in video summarization,” *IEEE Trans. Multimedia*, vol. 7, no. 5, pp. 907–919, Oct. 2005.

[53] C. Kanan, M. H. Tong, L. Zhang, and G. W. Cottrell, “SUN: Top–down saliency using natural statistics,” *Vis. Cogn.*, vol. 17, no. 6, pp. 979–1003, 2009.

[54] S. Frintrop, E. Rome, and H. I. Christensen, “Computational visual attention systems and their cognitive foundations: A survey,” *ACM Trans. Appl. Percept.*, vol. 7, no. 1, pp. 1–46, 2010.

[55] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[56] D. Walther and C. Koch, “Modeling attention to salient proto-objects,” *Neural Netw.*, vol. 19, no. 9, pp. 1395–1407, 2006.

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

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

[59] D. M. Chandler and S. S. Hemami, “VSNR: A wavelet-based visual signal-to-noise ratio for natural images,” *IEEE Trans. Image Processing*, vol. 16, no. 9, pp. 2284–2298, Sept. 2007.

[60] M. Barkowsky, J. Bialkowski, B. Eskofier, R. Bitto, and A. Kaup, “Temporal trajectory aware video quality measure,” *IEEE J. Select. Topics Signal Processing*, vol. 3, no. 2, pp. 266–279, 2009.

[61] U. Engelke and H.-J. Zepernick, “Psychophysical assessment of perceived interest in natural images: The ROI-D database,” in *Proc. SPIE/IEEE Int. Conf. Visual Communications and Image Processing*, Dec. 2011.

[62] U. Engelke, A. J. Maeder, and H.-J. Zepernick, “Visual attention modelling for subjective image quality databases,” in *Proc. IEEE Int. Workshop Multimedia Signal Processing*, Oct. 2009, pp. 1–6.

[63] R. J. Peters, A. Iyer, L. Itti, and C. Koch, “Components of bottom–up gaze allocation in natural images,” *Vision Res.*, vol. 45, no. 18, pp. 2397–2416, Aug. 2005.

[64] C. M. Masciocchi, S. Mihalas, D. Parkhurst, and E. Niebur, “Everyone knows what is interesting: Salient locations which should be fixated,” *J. Vision*, vol. 9, no. 11:25, pp. 1–22, 2009.

[65] J. Wang, D. M. Chandler, and P. Le Callet, “Quantifying the relationship between visual salience and visual importance,” in *Proc. IS&T/SPIE Human Vision and Electronic Imaging XV*, Jan. 2010, vol. 7527.

[66] M. S. Castelhano, M. L. Mack, and J. M. Henderson, “Viewing task influences eye movement control during active scene perception,” *J. Vision*, vol. 9, no. 3:6, pp. 1–15, 2009.

[67] 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 Southwest Symp. Image Analysis and Interpretation*, Mar. 2008, pp. 73–76.

[68] U. Engelke, “Modelling perceptual quality and visual saliency for image and video communications,” *Ph.D. dissertation, Blekinge Inst. Technol., Karlskrona*, Sweden, 2010.

[69] O. Le Meur, A. Ninassi, P. Le Callet, and D. Barba, “Do video coding impairments disturb the visual attention deployment?” *Signal Process. Image Commun.*, vol. 25, no. 8, pp. 597–609, 2010.

[70] J. S. Lee, F. de Simone, and T. Ebrahimi, “Influence of audio-visual attention on perceived quality of standard definition multimedia content,” in *Proc. Int. Workshop Quality Multimedia Experience*, July 2009, pp. 13–18.