

IMPROVING IMAGE QUALITY ASSESSMENT WITH MODELING VISUAL ATTENTION

Junyong You¹, Andrew Perki¹, Moncef Gabbouj²

1 Centre for Quantifiable Quality of Service in Communication Systems (Q2S)*, Norwegian University of Science and Technology, Trondheim, Norway;
2 Tampere University of Technology, Tampere, Finland

ABSTRACT

Visual attention is an important attribute of the human visual system (HVS), while it has not been explored in image quality assessment adequately. This paper investigates the capabilities of visual attention models for image quality assessment in different scenarios: two-dimensional images, stereoscopic images, and Digital Cinema setup. Three bottom-up attention models are employed to detect attention regions and find fixation points from an image and compute respective attention maps. Different approaches for integrating the visual attention models into several image quality metrics are evaluated with respect to three different image quality data sets. Experimental results demonstrate that visual attention is a positive factor that can not be ignored in improving the performance of image quality metrics in perceptual quality assessment.

Index Terms— Visual attention, saliency, fixation, image quality metric

1. INTRODUCTION

Perceptual image quality assessment plays an important role in digital image technology, such as the development and optimization of image compression and transmission schemes. Subjective quality assessment is considered to be the most reliable way to evaluate the quality of image presentations, but it is time-consuming. Over the years, a number of researchers have contributed significant research in the design of image quality assessment algorithms, claiming to have made headway in their respective domains [1]. According to the availability of reference image, image quality metrics (IQM) can be classified into three categories: full reference, reduced reference, and no reference. Most image quality metrics take into account the attributes of the human visual system (HVS), e.g. some distortions can not be perceived by human eyes because of the contrast

sensitivity function (CSF) [2], human vision is adapted to extract the structural information from a field of view [3], etc.

Visual attention is an important feature of the HVS. Many psychological and physiological experiments have demonstrated that human attention is not allocated equally to all regions in the field of view, but focused on certain attention regions [4]. Similarly to subjective image quality assessment, the most reliable method to detect attention regions is to use an external device, such as an eye-tracking device, to track human eye movements when viewing a scene. Eye movement is typically divided into fixation and saccade. Fixation is the maintaining of the visual gaze on a single location. Saccade refers to a rapid eye movement. Human do not look at a scene in fixed steadiness, instead, the human fovea sees only a small region (the central 2° of visual angle) in a field of view and fixes on this target, then moves to another target by saccadic eye movement [5]. However, eye-tracking is also time-consuming and cannot be performed in real-time applications. Thus, some researchers have tried to detect attention regions and find eye fixations from a field of view using computable and automatic approaches based on low-level visual features [4][6].

Although objective image quality assessment and visual attention analysis have been studied widely, few studies have been done on integrating visual attention into image quality assessment. Does visual attention analysis have a positive impact on visual quality assessment? Different experiments seem to give different conclusions. Ninassi et al. [7] observed that integrating visual attention into spatial pooling schemes in image quality assessment is not always advantage based on their eye-tracking experiments. In our previous works, we also found that the performance improvement on image quality assessment using the Saliency model in [4] is marginally [8], and even a saliency attention based spatial pooling scheme has a negative impact on video quality assessment for packet loss streams [9]. However, Liu et al. [10] reported that visual attention can be beneficial for two objective image quality metrics based on “ground truth” visual attention data from an eye-tracking experiment on natural images. We also found that visual attention detection can improve the accuracy in predicting

*“Centre for Quantifiable Quality of Service in Communication Systems, Centre of Excellence” appointed by the Research Council of Norway, funded by the Research Council, NTNU and UNINETT.

video quality with general degradation [8] and image quality assessment in Digital Cinema (DC) scenario [11].

In this work, we attempted to investigate the capability of visual attention in image quality assessment in three different scenarios: two-dimensional (2D) image, stereoscopic (3D) image, and Digital Cinema setup. Two different visual attention models [4][6] and an improved model based on the Saliency model with appropriate adjustments, such that they can be integrated into image quality assessment, are combined with several image quality metrics in order to evaluate the performance gain with respect to subjective image quality tests.

The remainder of this paper is organized as follows. Section 2 introduces the employed visual attention models. Five image quality metrics and the combination approaches with the visual attention models are presented in Section 3. Experimental results and discussions are presented in Section 4, and finally, some concluding remarks are given in Section 5.

2. VISUAL ATTENTION MODELS

The studies of visual attention model can be divided into two categories: top-down and bottom-up approaches. The top-down approach is usually driven by a certain task when viewing a scene, such as searching for a specific target from a field of view. Thus, top-down attention models are usually built based on visual features which are correlated with such task. In the bottom-up approach, a computational model for detecting visual attention regions is constructed based on low-level features of visual signals. In this study, we employed two bottom-up attention models: Saliency model [4] and GAFFE model [6].

Inspired by the behavior and the neuronal architectures of the early primate visual system, the Saliency model is to construct a single topographical saliency map first by combining multi-scale image features, such as colors, intensity, orientations and other visual information. Then, a winner-take-all network that implements a neurally distributed maximum detector is performed to detect the most salient locations step by step [4]. In the meanwhile, a saliency map of an image can be computed that can depict the saliency distribution over different locations.

According to a statistical analysis on image features at observers' gaze in eye-tracking experiments, Rajashekar et al. [6] proposed a GAFFE model that can find fixations in

Table I Descriptions of IQMs

Metrics	Description
PSNR	Peak signal-to-noise ratio
UQI	Universal quality index
SSIM	Single scale structural similarity measure
PHVS	Modified PSNR based on HVS
JND	Just noticeable distortion model

an image one by one based on four foveated low-level image features: luminance, contrast, and band-pass outputs of both luminance and contrast. It was found that image patches around human fixations have higher values of each of these features than image patches selected at random. Therefore, four respective saliency maps can be calculated based on these features on a foveated image, and they are combined linearly into a new map. The current fixation can be determined by choosing the most salient point on the new map.

In this work, we employed the Saliency model to compute the saliency map based on four low-level features (color, intensity, orientation, and skin), the GAFFE model to generate the combined map, and another attention model, called saliency attention model. We have found that contrast information is an important factor in image quality assessment [11]. Human usually pay more attention to those regions that have higher contrast levels. Thus, an image can be divided into different blocks, and the standard deviation of each block was used to denote the contrast information of this block. Additionally, since human usually pay more attention to the regions close to the image center, we can use a normalized Gaussian filter (G) with the center located at the image center to assign a weight to the position of image regions. Finally, the saliency attention model is to combine the saliency map (S), the contrast map (C), and the Gaussian filter to generate another map, called saliency attention map, as in Eq. (1). Figure 1 gives an illustration of an image and its different attention maps.

$$A = G \cdot (S + 0.5 \cdot C) \quad (1)$$

3. VISUAL ATTENTION MODELS INTEGRATED IMAGE QUALITY METRICS

3.1. Objective image quality metrics

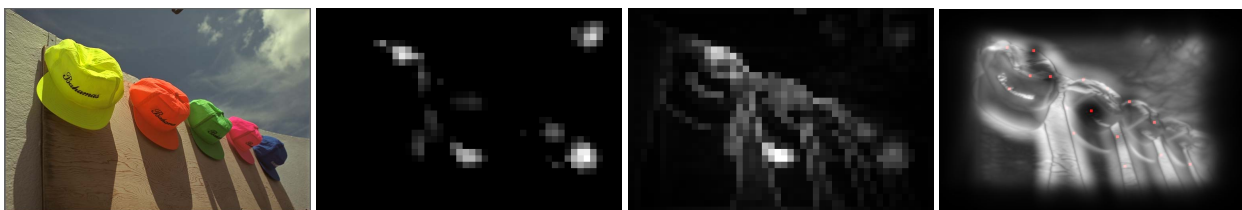


Fig. 1. Illustration of visual attention maps (red points in (d) denote the fixations detected by the GAFFE model).

In order to investigate the capability of visual attention in image quality assessment, five objective IQMs, summarized in Table I, were employed in this work. Different attributes of the HVS are applied in these metrics, more or less. For example, UQI [12] and SSIM [3] are constructed based on the conclusion that human vision is more sensitive to the degradation on structural information of an image, such as the distortions in edge regions. Due to the fact that human eyes cannot perceive all distortions, PHVS [2] and JND [13] models are established based on the CSF and masking effect of the HVS from the Discrete Cosine Transform (DCT) domain and pixel domain, respectively.

Although these IQMs are constructed on different principles, they have a common ground – a quality map (PSNR is based on MSE map), which can depict local distortions at each pixel or in each small region, is calculated first for each metric and then the overall image quality is taken as the average or based on the average of this quality map over all pixels/regions. Therefore, it is possible to integrate visual attention models into the quality map with different approaches.

3.2. Integrating visual attention into IQMs

As mentioned earlier, an attention map that can depict the distribution and intensity of visual attention regions in an image can be computed using individual attention models (Saliency model, GAFFE model, and saliency attention model), and a quality map between the reference and distorted images is calculated using the IQMs. In this study, the attention map was computed from a reference image. Subsequently, we will integrate the attention map into the quality map in different approaches. The first approach is to combine the attention map, the quality map, and the computed attention map values. For example, let the attention map be A whose element depicts the intensity of attention regions at each pixel or in each small region, and the quality map be Q , these two maps can be combined according to the following functions:

$$IQ = \begin{cases} avg(A \cdot Q) \\ avg(sub_A \cdot Q) \end{cases} \quad (2)$$

where avg is the averaging operation over all pixels/regions, and sub_A denotes that only a subset of the attention regions is used. The attention map values were utilized to select the subset, and those regions with higher attention map values were selected to be participated in quality computation while other regions were excluded from the computation. In this study, we tested different ratios of a subset regarding the entire map and the result when the ratio was set to 15% is reported in the next section. In this approach, we can evaluate the capabilities of attention regions and the corresponding map values for image quality assessment.

The second approach is to integrate attention regions into the quality map without using the attention map values. On an attention map, the map values in those non-attention

regions are usually equal to 0, and the map values in attention regions were set to 1 in stead of the original map values. The overall image quality is then computed based on the adjusted attention map and the quality map using Eq. (2) again. In the second approach, the capabilities of attention regions for image quality assessment can be evaluated.

Finally, because the Saliency model and the saliency attention model are based on the unit of image blocks, we can extract image blocks which can attract more attention according to saliency map values or saliency attention map values. Therefore, the third approach is to divide an image into different blocks, and a quality value in an individual block is calculated using IQMs first. These quality values are then weighted by the attention map values and averaged over all blocks to obtain an overall quality of an image. Similarly to the above two approaches, the entire attention map and a subset are also evaluated, respectively, in the third approach. This approach is supposed to study the relation among visual attention, local block quality and overall image quality, since human eyes usually cannot see an entire image at once, especially in Digital Cinema setup, due to the fact that the visual acuity angle is about 2° only. In the third approach for the GAFFE model, we chose image blocks whose middles were the fixations detected by the GAFFE model, and image quality was computed as an average over these blocks. The block size (S) was calculated as following:

$$S = 2 \cdot 6 \cdot H \cdot \tan\left(\frac{1}{180}\right) \quad (3)$$

where H denotes the height of an image, since the visual acuity angle is about 2° and the viewing distance is usually set to $6H$ in subjective image quality assessments.

4. EXPERIMENTS AND DISCUSSION

In this work, we attempt to investigate the capability of visual attention in image quality assessment in three different scenarios: 2D images, 3D images, and DC. Three respective datasets were employed, including LIVE image quality database [14], stereoscopic image quality database [15], and an image quality dataset in DC setup [11].

Because different subjective quality assessment may use different quality scales, a nonlinear regression operation between predicted image quality values (IQ) and the subjective scores (MOS or DMOS) was performed by a logistic function in Eq. (4), as suggested in a Video Quality Experts Group (VQEG) report [16].

$$MOS_p = \frac{a_1}{1 + \exp[-a_2 \cdot (IQ - a_3)]} \quad (4)$$

where a_1 , a_2 , and a_3 denote the repression parameters. The nonlinear regression function was used to transform the set of metric values to a set of predicted (D)MOS values, MOS_p , which were compared against the actual subjective scores and then resulted in a criterion: Pearson correlation coefficient (PCC).

Table II *PCC* of original IQMs on different image quality datasets

Data sets	PSNR	UQI	SSIM	PHVS	JND
LIVE	0.814	0.859	0.864	0.878	0.815
3D	0.795	0.825	0.677	0.769	0.738
DC	0.914	0.860	0.888	0.938	0.925

Table III Relative gain (%) of *PCC* in visual attention based image quality assessment

Attention models			PSNR	UQI	SSIM	PHVS	JND
Saliency model	First approach	Entire map	2.3/0.8/2.0	1.4/2.8/-1.3	0.5/5.0/7.0	1.3/2.3/0.6	1.6/7.7/2.6
		Subset	2.1/0.8/2.0	1.4/2.8/-1.3	0.5/5.0/7.0	1.4/2.3/0.6	1.6/7.7/2.6
	Second approach	Entire map	3.3/0.5/1.0	2.9/3.5/6.3	2.0/5.8/0.9	2.0/2.5/1.2	2.4/7.3/1.4
		Subset	0.9/1.0/0	1.4/0/-0.2	1.5/0/0	1.7/0.3/0	0.5/7.3/0
	Third approach	Entire map	3.2/0.7/2.0	2.2/1.9/7.8	-1.5/6.5/5.2	-0.2/2.5/2.6	1.8/1.0/2.9
		Subset	3.9/-0.8/4.0	4.0/-2.1/5.6	1.2/10.8/8.6	1.5/1.9/3.0	2.0/-0.2/3.0
Saliency attention model	First approach	Entire map	3.6/-0.3/3.4	4.8/-0.3/3.5	3.8/9.5/8.6	3.6/4.5/2.1	3.0/7.0/2.7
		Subset	3.9/-0.8/4.6	2.9/-0.4/8.3	4.2/10.6/9.0	3.9/5.0/1.7	3.4/6.5/3.6
	Second approach	Entire map	0.1/1.0/0	0/0/-0.1	0.1/0/0	0.5/0.1/-0.2	0.1/7.3/0
		Subset	3.8/0.4/4.3	4.8/-0.4/5.8	3.4/12.9/9.0	3.7/1.2/1.9	3.4/7.7/3.1
	Third approach	Entire map	3.8/0.3/4.6	3.5/-0.8/9.9	1.7/12.3/8.6	2.8/1.3/3.4	2.4/0.6/3.6
		Subset	3.6/1.0/4.2	4.2/0.3/10.7	2.2/12.2/9.2	3.8/1.3/3.8	2.7/1.3/4.1
GAFFE model	First approach	Entire map	1.0/0.9/2.2	0.8/-0.4/2.1	1.4/1.8/6.3	1.2/1.7/0.9	1.0/2.2/1.0
		Subset	2.8/-0.5/3.0	2.8/-0.4/-1.4	2.9/8.2/8.5	2.0/1.1/1.8	2.4/1.2/1.2
	Second approach	Entire map	0/0/0.1	0.1/0/0	0.4/0/0	0.2/0/-0.1	0/-0.1/0
		Subset	2.0/-0.5/2.3	2.0/-0.4/-1.8	2.7/10.9/8.1	1.9/0.7/2.0	1.8/-0.9/2.0
	Third approach	Fixation blocks	4.0/-1.0/4.6	3.8/-5.7/5.0	3.1/9.5/9.3	2.4/-6.8/3.3	3.5/-3.1/2.6

Table II gives the evaluation results of the original metrics on three image quality datasets in terms of *PCC*, and Table III reports the relative gain of *PCC* in different approaches, in which the first value in a cell denotes the *PCC* gain in 2D image quality dataset, the second value is for 3D dataset, and the last value is for DC dataset. The relative gain of *PCC* is calculated as following:

$$\text{gain} = \frac{PCC_{\text{attention}} - PCC_{\text{original}}}{PCC_{\text{original}}} (\%) \quad (5)$$

where $PCC_{\text{attention}}$ denotes the Pearson correlation coefficient on an individual quality dataset using a certain attention integrated IQM, and PCC_{original} is the correlation of the original metrics. Figure 2 illustrates the bar charts of *PCC* values when integrating the saliency attention model into image quality assessment in different scenarios.

Statistically speaking, visual attention models can improve the performance of IQMs according to the

experimental results. Most exceptions appear in stereoscopic image quality assessment. In our opinion, the reason is that stereoscopic image quality is not only determined by image content and human attention, but also other stereopsis attributes, such as depth information. In addition, the employed visual attention models are both for two-dimensional vision, while a special attention model to detect attention regions from 3D scenes might be required for stereoscopic image quality assessment.

Based on a statistic analysis on the experimental results, the saliency attention model has the strongest capability in improving the performance of IQMs, in turn are the Saliency model and the GAFFE model. The first approach that used attention map values as weights is statistically better, even not very evidently, than the second approach in which attention map values were not taken into account in computing image quality. This observation demonstrates that attention map values, which can depict attentive

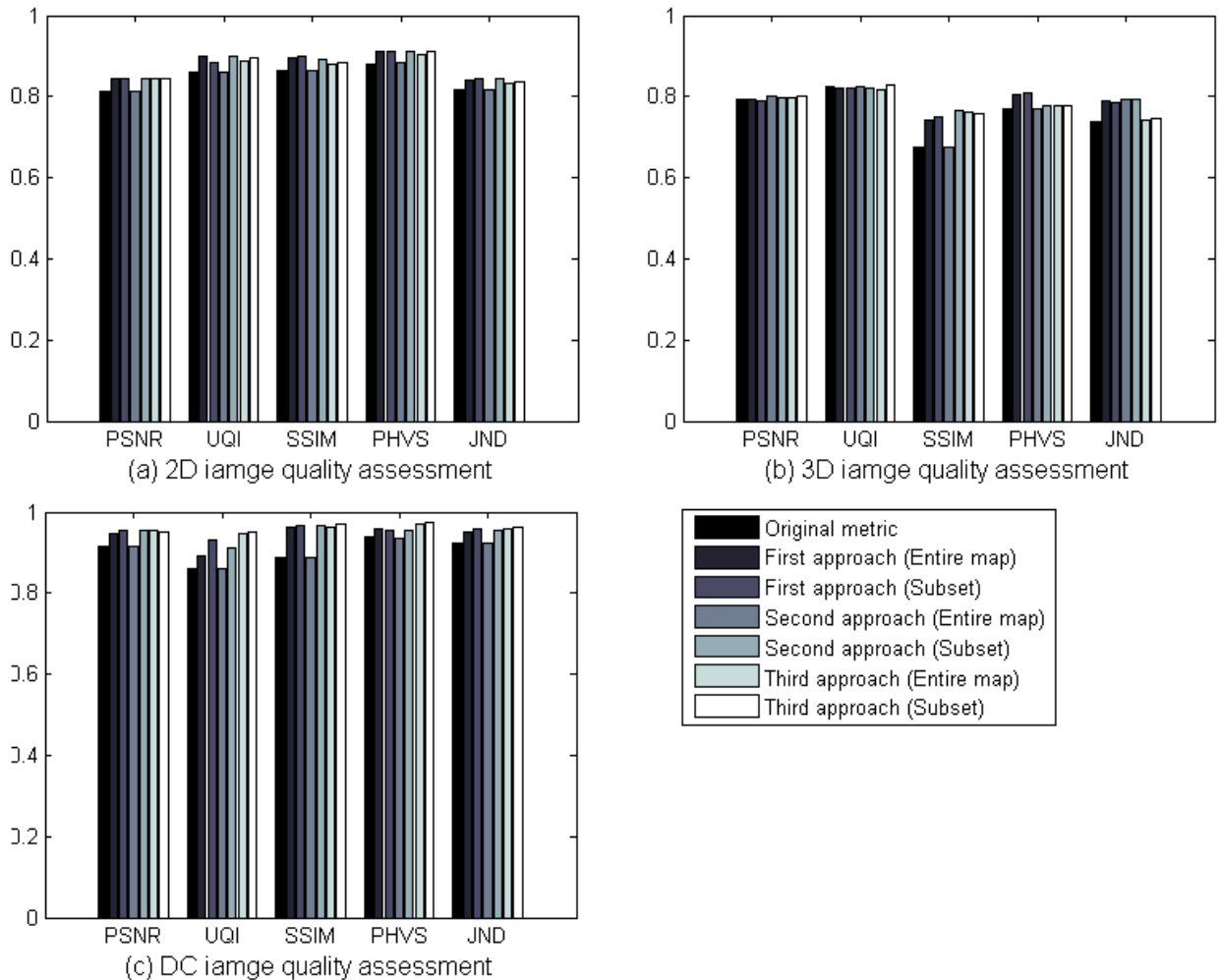


Fig. 2. Bar charts of PCC values of IQMs in different scenarios (The first bar for each metric denotes the *PCC* value of the original metric, and in turn are entire map and subset in the first approach, and the second and the third approaches).

intensity over different regions, might be a beneficial factor for improving the performance of IQMs. The third approach, based on local block quality, is better than the first and the second approaches in most cases, especially for Digital Cinema setup. We believe the reason is that Digital Cinema always has a large size screen, so that subjects can not see an entire image at once [11]. Compared to the use of an entire attention map, using a subset of the attention map usually has better performance, except for the Saliency model using the second approach. In this work, we found that the best performance is always achieved when 10-20% image blocks regarding an entire image were participated in quality computation. Similarly, the third approach for the GAFFE model is better than using original metrics, except for 3D scenario. This observation partially confirms our conclusions in [9] that the perceived visual quality is usually determined by some certain regions that have particular characteristics, such as attentive intensity and severe distortion, rather than an entire image.

Although visual attention can improve the performance of IQMs, the computation cost for running attention models

is an important issue, especially in real-time applications. Taking an example of SSIM and the Saliency model in the LIVE dataset, the computation time of the Saliency model is approximately 4 times longer than running SSIM. However, as the attention models were performed on a reference image, the computation costs for attention models can be shared equally on all distorted images if a reference image has many distorted presentations. Furthermore, the results of attention models might be affected by distortion information in a distorted image. For example, it has been found that compression-type distortions (JPEG, JPEG 2000, packet losses) in which the distortions are spatially localized can change viewers' focus in images as well as fixation durations, while white noise distributed over the entire image has almost no influence on visual attention [17]. Thus, in future work, we will investigate the performance of visual attention models performed on not only original undistorted images, but also distorted images. Additionally, the task of quality assessment has also a significant effect on eye movement [18]. Hence, although the employed bottom-up attention models have showed a promising performance,

we believe that a top-down attention model might be more suitable for image quality assessment. This issue needs further investigation in future work.

5. CONCLUSIONS

In this paper, we have evaluated the capabilities of visual attention models in image quality assessment by integrating them into IQMs in different approaches for three scenarios. Three bottom-up attention models and several image quality metrics were employed. The experimental results demonstrated that visual attention is an important factor in evaluating the perceived image quality. In the future, we will investigate more appropriate methods to combine visual attention models and image quality metrics, and the performance of attention models when they are performed on distorted image presentations. Furthermore, top-down attention models driven by the task of quality assessment will also be taken into account in developing attention based visual quality metrics.

REFERENCES

- [1] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms," *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3441-3452, Nov. 2006.
- [2] N. Ponomarenko, F. Battisti, K. Egiazarian, et al., "On Between-coefficient Contrast Masking of DCT Basis Functions," in *Proc. Int. Workshop Video Processing and Quality Metrics*, Scottsdale, Arizona, USA, Jan. 2007.
- [3] Z. Wang, A. C. Bovik, H. R. Sheikh, et al., "Image Quality Assessment: from Error Visibility to Structural Similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600-612, Apr. 2004.
- [4] L. Itti, and C. Koch, "Computational Modeling of Visual Attention," *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194-203, Mar. 2001.
- [5] G. A. Carpenter, "Movements of the eyes," London: Pion, 1988.
- [6] U. Rajashekar, I. V. D. Linde, A. C. Bovik, et al. "GAFFE: A Gaze-Attentive Fixation Finding Engine," *IEEE Trans. Image Processing*, vol. 17, no. 4, pp. 564-573, Apr. 2008.
- [7] A. Ninassi, O. L. Meur, P. L. Callet, and D. Barba, "Does Where You Gaze on an Image Affect Your Perception on Quality? Applying Visual Attention on Image Quality Metric," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, pp. II 169-172, San Antonio, Texas, USA, Sep. 2007.
- [8] J. You, A. Perkis, M. M. Hannuksela, and M. Gabbouj, "Perceptual Quality Assessment based on Visual Attention Analysis," in *Proc. ACM Int. Conf. Multimedia*, pp. 561-564, Beijing, China, Oct. 2009.
- [9] J. You, J. Korhonen, and A. Perkis, "Spatial and Temporal Pooling of Image Quality Metrics for Perceptual Video Quality Assessment on Packet Loss Streams," *IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 1002-1005, Dallas, USA, Mar. 2010.
- [10] H. Liu, and I. Heyderichx, "Studying the Added Value of Visual Attention in Objective Image Quality Metrics Based on Eye Movement Data," in *Proc. IEEE Int. Conf. Image Processing*, pp. 3097-3100, Cairo, Egypt, Nov. 2009.
- [11] J. You, F. N. Rahayu, U. Reiter, and A. Perkis, "HVS-based Image Quality Assessment for Digital Cinema," in *Proc. SPIE Image Quality and System Performance VII*, vol. 7529, San Jose, USA, Jan. 2010.
- [12] Z. Wang, and A. C. Bovik, "A Universal Image Quality Index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81-84, Mar. 2002.
- [13] X. K. Yang, W. S. Ling, Z. K. Lu, et al., "Just Noticeable Distortion Model and Its Applications in Video Coding," *Signal Processing: Image Communication*, vol. 20, no. 7, pp. 662-680, Aug. 2005.
- [14] H. R. Sheikh, Z. Wang, et al., "LIVE Image Quality Assessment Database," <http://live.ece.utexas.edu/research/quality>.
- [15] J. You, L. Xing, A. Perkis, et al., "Perceptual Quality Assessment for Stereoscopic Images Based on 2D Image Quality Metrics and Disparity Analysis," in *Proc. Int. Workshop Video Processing and Quality Metrics*, Scottsdale, Arizona, USA, 2010.
- [16] VQEG, "Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II (FR-TV 2)," VQEG, Aug. 2003.
- [17] C. T. Vu, E. C. Larson, and D. M. Chandler, "Visual Fixation Pattern when Judging Image Quality: Effects of Distortion Type, Amount, and Subject Experience," *IEEE Southwest Symposium on Image Analysis and Interpretation*, pp. 73-76, Mar. 2008.
- [18] A. Ninassi, O. Le Meur, P. Le Callet, et al. "Task Impact on the Visual Attention in Subjective Image Quality Assessment," *European Conf. Signal Processing*, Florence, Italy, Sep. 2006.