A Perceptual Image Sharpness Metric Based on Local Edge Gradient Analysis

Christoph Feichtenhofer, Hannes Fassold and Peter Schallauer

***Abstract*—In this paper, a no-reference perceptual sharpness metric based on a statistical analysis of local edge gradients is presented. The method takes properties of the human visual system into account. Based on perceptual properties, a relationship between the extracted statistical features and the metric score is established to form a Perceptual Sharpness Index (PSI). A comparison with state-of-the-art metrics shows that the proposed method correlates highly with human perception and exhibits low computational complexity. In contrast to existing metrics, the PSI performs well for a wide range of blurriness and shows a high degree of invariance for different image contents.**

***Index Terms*—Image sharpness, image blur, no-reference, sharpness metric, perceptual, image quality assessment.**

I. Introduction

THE Human Visual System (HVS) provides the most reliable assessments in vision-based quality inspection, but in practice a subjective approach is costly and time-consuming. It is obvious that objective techniques are needed to predict the quality automatically by emulating human subjectivity.

Image quality metrics can be categorized by the amount of information available from the original undistorted image. Full-reference metrics have access to both the distorted image and the original image. Practically, the reference signal is unavailable in most applications. Thus, no-reference metrics, which rely only on the distorted image, are required. Various distortion types may occur during acquisition, processing, transmission and storage of digital content. No-reference metrics are typically designed for specific types of distortions. This work aims at sharpness estimation. Besides quality assessment, some further applications for sharpness metrics are for example automatic image enhancement [1], blind image deconvolution [2], camera auto-focus systems, the depth estimation of a scene [3], or image super-resolution [4].

Several spatial, as well as transform-based sharpness metrics are proposed and discussed in the literature; a comprehensive overview is given in [5] and [6]. Most of the spatial domain methods analyze regions around image edges, which are affected by a smoothing or smearing effect under blurring. Ferzli and Karam [5] and, more recently, Narvekar and Karam [7] propose spatial domain sharpness metrics, based on a concept of "just-noticeable blur". Their algorithms build on measuring the edge spreads, initially proposed by Marziliano *et al.* [8]. A spatial metric which does not concentrate on edge widths is proposed by Li *et al.* [9], where sample statistics in the vicinity of edges in the original and a re-blurred image are compared. Transform-based methods typically exploit the fact, that sharp images contain more high frequency content than blurred images (e.g [1], [10]). A sharpness metric based on the local phase coherence of complex wavelet coefficients is proposed in [11], where Hassen *et al.* utilize the fact that blur causes a disruption of local phase near sharp image features. A hybrid approach, combining spatial and transform-based features, is presented by Vu *et al.* [6]. They use the slope of the local magnitude spectrum and the local total variation to produce a sharpness map. The usually better performance of a hybrid approach comes with the cost of higher computational complexity. Recently, a transform-based method with very low computational complexity was proposed by Vu and Chandler [10], who measure the log-energy in high frequency discrete wavelet transform subbands.

Our method is also motivated by the basic idea of measuring sharpness around edges [5], [7], [8]; however, recent investigations [6], [10], [11] have shown that spatial-based metrics are outperformed by transform-based and hybrid metrics. In this work, we develop a metric, termed Perceptual Sharpness Index (PSI), that represents the perceived sharpness in an image. Our spatial-based PSI is able to consistently outperforms transform-based and hybrid metrics by using an improved way of computing edge widths, and novel approaches for the selection of relevant measuring points and the pooling of the local gradient features. We also analyze the edge slopes to integrate an acutance measure for modeling the influence of local contrast information on the perception of sharpness. Our algorithm is computationally fast and requires no training; all parameters are either based on properties of the HVS, or determined from our implementation study on the perceptual performance of the PSI.

II. Perceptual Sharpness Index (PSI)

A. Adaptive Edge Selection

The first step of the algorithm is the generation of an edge map, representing the relevant edges in the image. To this end, a vertical and horizontal Sobel filter is applied to the luminance component of the image, followed by a thresholding and thinning process. The Sobel filter responses are pointwise squared and summed to form a measure of the squared image

C. Feichtenhofer is with the Institute of Electrical Measurement and Measurement Signal Processing, Graz University of Technology, Inffeldgasse 23/II, 8010 Graz, Austria. (e-mail: cfeichtenhofer@gmail.com).

H. Fassold and P. Schallauer are with the Institute for Information and Communication Technologies, Joanneum Research Digital, Steyrergasse 17, 8010 Graz, Austria. (e-mail: hannes.fassold, peter.schallauer@joanneum.at).

gradient G, which is thresholded with an adaptive threshold

$$T = \alpha \overline{G}, \tag{1}$$

where $\overline{G}$ denotes the mean of G and the scaling factor α is set based on the experimental study on the influence of this parameter in Section III. By considering the average edge gradient energy, the adaptive threshold (1) leads to a focus on the most significant edges in the image. This focus on high local contrasts adapts to human perception, because the information in the HVS is represented in terms of contrast instead of the absolute intensities [12]. After thresholding, a thinning process is performed as a nonmaximum suppression along the edge gradients to generate an edge map.

B. Edge Width Measurement

We measure the widths of these significant edges $\mathbf{x} = (x, y)$ by pixel-wise tracing along the edge gradient. Only approximately vertical gradients are measured to ensure low computational complexity and reasonably accurate measurements; our own experiments, and tests in [8], show that considering additional horizontal measurements do not improve the performance. Surprisingly, we found that incorporating the measurements of diagonal edge gradients resulted in a performance decrease. We believe, this is due to the larger quantization step-width ($\sqrt{2}$) of diagonal pixels, compared to a spacing of one between pixels in horizontal/vertical direction.

The angle difference between the gradient direction $\phi(\mathbf{x}) = tan^{-1}\left(\frac{I_y(\mathbf{x})}{I_x(\mathbf{x})}\right)$ and the tracing direction is denoted by $\Delta\phi(\mathbf{x})$, where the image gradient components I_x and I_y are computed by using finite differences. Consequently, at every edge $\mathbf{x}$, with approximately vertical gradient direction $\Delta\phi(\mathbf{x}) < \Delta\phi_{max}$, the edge width is computed by

$$w(\mathbf{x}) = \frac{w_{up}(\mathbf{x}) + w_{down}(\mathbf{x})}{cos(\Delta\phi(\mathbf{x}))}, \tag{2}$$

where $w_{up}(\mathbf{x})$ and $w_{down}(\mathbf{x})$ are the distances between the detected edge pixel $\mathbf{x}$ and the traced local maximum $I_{max}(\mathbf{x})$ and local minimum $I_{min}(\mathbf{x})$ luminance pixels, respectively. Although the denominator in eq. (2) compensates for deviations $\Delta\phi(\mathbf{x})$ of the edge gradient direction to the tracing direction, large deviations will incorporate errors into the measurement process. Thus, only approximately vertical edge gradients, with a maximum angle difference of $\Delta\phi_{max}$, are considered. Additionally, a measurement point is rejected if the corresponding gradient starts or ends at an image border.

C. Modeling the Human Perception of Acutance

Receptive fields of retinal cells are much more sensitive to contrast across a border than to the overall light level [12]. Therefore, the HVS is more sensitive to signal contrast than to the local absolute signal strength. Motivated by the observation that high-contrast edges are perceived as sharper, we refine the metric by taking the edge slopes explicitly into account. The slope is utilized as an acutance measure, and is given by:

$$m(\mathbf{x}) = \frac{I_{max}(\mathbf{x}) - I_{min}(\mathbf{x})}{w(\mathbf{x})}, \tag{3}$$

assuming a luminance range of $[0, 1]$. The measured edge width is decreased in proportion to the edge slope, to assess high contrast edges as sharper:

$$w_{\text{PSI}}(\mathbf{x}) = \begin{cases} w(\mathbf{x}) - m(\mathbf{x}), & \text{if } w(\mathbf{x}) \geq w_{\text{JNB}}, \\ w(\mathbf{x}) & \text{otherwise.} \end{cases} \tag{4}$$

To imply the just-noticeable difference, perceived by the HVS, only widths above a minimum width w_{JNB} are reduced. This minimum width is the Just-Noticeable Blur (JNB) width, representing the minimum amount of perceived blurriness around an edge, which depends on the local contrast in the neighborhood of the edge. It is estimated in subjective experiments [5] on a blurred edge with different blur amounts, to 5 pixels for edges with a contrast lower than 50, and 3 pixels for edges with higher contrast. Since the proposed PSI only focuses on high contrast edges we use $w_{\text{JNB}} = 3$.

D. Block Based Percentile Pooling

1) Local Sharpness Estimation: To deduce local sharpness values, the image is divided into blocks of 32×32 pixels. Due to the quantization step of one pixel in the edge width measurement process, narrow edges may generate a high quantization error. Therefore, we require at least one measurement from wide edges, or more than one measurements from narrow edges in each block. Essentially, this corresponds to a minimum sum of 2 of all measured widths within a block. We then compute the local sharpness estimates as the reciprocal of the average $w_{\text{PSI}}(\mathbf{x})$ of all measurements in a local block.

2) Global Sharpness Estimation: To ignore out of focus regions in images with strong depth-of-field effects, and to adapt to human sharpness perception, the PSI is calculated by a percentile pooling of the local sharpness estimates as the highest γ^{th} percentile average of the local sharpness values across the image. This approach has proven to correlate well with subjective perception [13], [14]. The motivation behind the percentile pooling is the observation that the sharpest regions in an image heavily influence human perception of sharpness.

III. Perceptual Implementation Analysis

To determine the influence of the parameters, which do not base on subjective experiments (*i.e.* $\Delta\phi_{max}$, α and γ), we apply the PSI to the Gaussian blurred images from the subjective quality assessment database IVC [15] and vary these parameters. We also provide the optimal parameters[1] for the PSI by analyzing its performance on all publicly available subjective databases with Gaussian blurred images (*i.e.* LIVE [16], TID2008 [17], CSIQ [18] and IVC). Nevertheless, our evaluations in Section IV show results for the parameters[2] based on IVC only, since using the optimal parameters would imply a training on the test data; however, when applying the PSI to arbitrary images the optimal parameters[1] should be chosen.

We investigate the impact of all combinations of the three parameters. First, we observe that varying the maximum angle difference has only modest effect on performance; generally, low thresholds $\Delta\phi_{max} \leq 2^\circ$ decrease performance, as they

[1] $\Delta\phi_{max} = 7^\circ; \alpha = 5.6; \gamma = 18$ (optimal for LIVE, TID, CSIQ and IVC)
[2] $\Delta\phi_{max} = 8^\circ; \alpha = 4.7; \gamma = 22$ (used for all evaluations in this paper)

do not incorporate enough measurements, and larger thresholds $\Delta\phi_{max} \geq 15°$ also decreases performance, due to the integration of imprecise measurements. Second, we investigate the impact of multiple scaling factors α for the adaptive threshold, which focuses the PSI on high local contrasts. Scaling factors $\alpha > 3$ increase the overall performance, due to the focus on high contrast edges, which exhibit higher visual importance; however, using factors $\alpha > 9$ turns out to decrease the performance, because not enough edge measurements are integrated. This also applies for very low percentiles $\gamma < 2$ for the pooling. However, reasonably low pooling percentages lead to a concentration on the sharpest image blocks; we found that percentiles $\gamma < 40$ are essential for good performance when assessing the sharpness in images with out of focus regions, *e.g.*, if the foreground is sharp and the background is blurred.

IV. Experimental Results

In this Section we evaluate the performance of the proposed PSI and compare it against five state-of-the-art metrics: two spatial sharpness metrics: JNB [5], CPBD [7], two spectral sharpness metrics: S_3 [6], FISH [10], and two no-reference quality metrics: DIIVINE [19] and BLIINDS-II [14], by using the implementations provided by the authors.

The capability of an image quality metric is typically judged by how well it correlates with human perception of quality (Section IV-A) and how suited it is for real-time applications (Section IV-B). A sharpness metric should additionally decrease monotonically as the blurriness of an image increases (Section IV-C), and exhibit small variations for different images with same blur level (Section IV-D).

A. No-Reference Quality Assessment

To predict the quality of several distorted images, we use the three largest subjective image quality databases LIVE, TID and CSIQ. These databases provide mean opinion scores (MOS) from people rating the image quality as ground truth. The Gaussian-blur datasets, consisting of 174, 150 and 100 images from LIVE, CSIQ and TID, respectively, are used.

We report the following standard measures to describe the performance of the predictions: *1*) Pearson's linear correlation coefficient (LCC) indicating the prediction accuracy; *2*) the Spearman rank-order correlation coefficient (SROCC), expressing the monotonicity by ignoring the relative distance between the data; *3*) the Outlier ratio (OR), indicating the consistency, is the fraction of predictions outside the interval [MOS - 2σ, MOS + 2σ], where σ is the standard deviation of the opinion scores for a single image; and *4*) the root-mean-squared error (RMSE). High LCC and SROCC, as well as low OR and RMSE indicate good predictions of the MOS.

We also show performance scores, conforming to the evaluation procedures suggested by the Video Quality Experts Group [20], where a non-linear mapping of the predictor results s_i to the subjective ratings is recommended; each metric score is first mapped, using the suggested four-parameter logistic function

$$\mathrm{MOS}(s_i) = \frac{\beta_1 - \beta_2}{1 + \exp\left(\frac{-(s_i - \beta_3)}{|\beta_4|}\right)} + \beta_2, \qquad (5)$$

where the parameters $\beta_1, \beta_2, \beta_3, \beta_4$ are chosen to minimize the sum of squared differences between the predicted s_i and the MOS. Note, however, that these scores show a less critical judgment due to the mapping to the desired MOS, also note that the SROCC does not change due to this mapping, as it only assesses the correct ordering.

Table I presents the performance for quality prediction of blurred images. The MOS scores are scaled differently in each database and therefore the RMSE diverges. Since DIIVINE and BLIINDS-II are trained on LIVE, their performance is not listed for this database. Overall, the PSI performs very well in terms of correlation with human perception, it is significantly outperforming the other metrics and the best performer or tied with the best performer on all databases. Only S_3 is competitive, but at much higher computational complexity (*i.e.* our PSI is around 400 times faster than S_3). Furthermore, DIIVINE and BLIINDS-II also perform well, even though these metrics are adjusted for the detection of four additional distortion types. To express the overall performance gain, we weight the LCC scores with the number of images in the respective databases. Our PSI improves the performance over the second best metric (S_3) by 11.31% and 2.30% in LCC and LCC after non-linear mapping (5), respectively.

TABLE I
Performance and Computational Complexity Comparison on the Blur Datasets, the Two Best Performers are Bolded

	Hybrid	Transform-based			Spatial-based		
	S_3	FISH	DIIVINE	BLIINDS-II	JNB	CPBD	PSI
LCC							
LIVE	0.8958	0.9096	trained	trained	0.8222	**0.9128**	**0.9535**
TID	0.7330	0.7430	**0.8137**	0.7813	0.6535	0.8046	**0.8366**
CSIQ	0.7703	**0.8516**	0.8493	0.8198	0.3416	0.8102	**0.8901**
SROCC							
LIVE	**0.9634**	0.9221	trained	trained	0.8423	0.9430	**0.9617**
TID	**0.8418**	0.7828	0.8237	0.8206	0.6667	0.8412	**0.8493**
CSIQ	**0.8860**	0.8703	0.8483	0.8471	0.7801	0.8644	**0.8897**
LCC after non-linear mapping (5)							
LIVE	**0.9189**	0.9178	trained	trained	0.8428	0.9128	**0.9535**
TID	**0.8541**	0.8079	0.8193	0.8260	0.6931	0.8235	**0.8546**
CSIQ	0.8836	**0.8890**	0.8625	0.8552	0.8261	0.8553	**0.9061**
OR after non-linear mapping (5)							
LIVE	**13.8%**	16.1%	trained	trained	23.6%	16.1%	**6.3%**
TID	**67.0%**	**67.0%**	74.0%	71.0%	73.0%	69.0%	**66.0%**
CSIQ	34.7%	**30.0%**	36.0%	31.3%	39.3%	38.0%	**26.0%**
RMSE after non-linear mapping (5)							
LIVE	**8.579**	8.6341	trained	trained	11.706	8.882	**6.557**
TID	**0.6104**	0.6915	0.6729	0.6614	0.8459	0.6657	**0.6094**
CSIQ	0.1342	**0.1312**	0.1450	0.1485	0.1615	0.1485	**0.1212**
Average speed (sec./image)							
LIVE	37.363	**0.0702**	29.305	113.034	0.7786	0.6927	**0.0783**
TID	20.335	**0.0327**	17.817	57.191	0.4242	0.3727	**0.0452**
CSIQ	24.994	**0.0611**	21.575	72.067	0.5416	0.5181	**0.0696**

B. Runtime Analysis

The average runtime for the compared algorithms is also listed in Table I. All metrics are implemented in MATLAB and executed on a 2.8 GHz processor with 4 GB RAM. The runtime scales with the image sizes which are typically 768×512, 512×384 and 512×512 for LIVE, TID and CSIQ, respectively. FISH is the fastest sharpness estimator, due to its special design for low computational complexity. The proposed PSI is also very fast and competitive to FISH, because it concentrates on a set of significant edges and therefore on only a small fraction of all pixels.

C. Monotonic Blur Prediction

In order to evaluate the metrics' capability to monotonically predict increasing amounts of blur, 1411 blurred versions of all reference images from the databases were generated by using a circular-symmetric 2D Gaussian kernel of standard deviation σ_{blur}, which ranges from 0.25 to 8.25, in discrete steps of 0.5 pixels. All of these images are natural, the TID database also contains an artificial image which is excluded in this experiment. The results for the four best performing metrics are visualized in Fig. 1, which shows the metric values for the blurred versions of the same reference image by individual lines plotted against the the standard deviation of the Gaussian blur kernel. The experiments reveal that all methods decrease monotonically when applied to identical images with increasing blurriness; however, for high amounts of blur only the proposed PSI is consistently decreasing. Moreover, the CPBD, FISH and S_3 metrics show a rapid decay for increasingly blurred content and are unable to appropriately quantify the amount of blur for heavily blurred images.

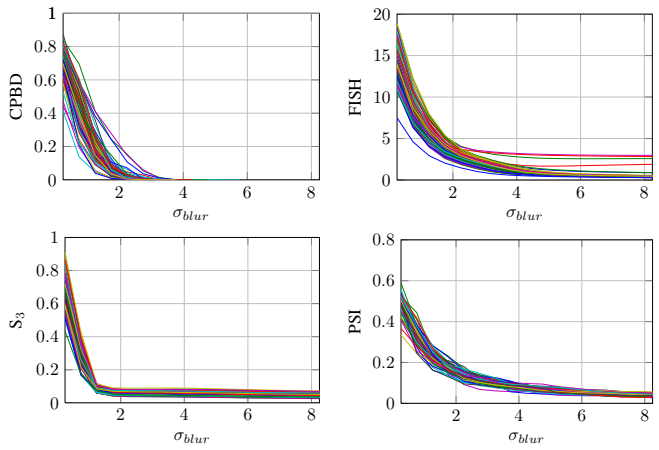

Fig. 1. Performance of different sharpness metrics on blurred versions of the 83 original images from LIVE, TID and CSIQ, each line represents an image, blurred with different blur amounts σ_{blur}. The proposed PSI shows a continuously monotonic decreasing behavior for increasing amount of blur.

D. Image Content Invariance

We also report the standard deviation of the sharpness metrics for slightly blurred versions of the 83 original images, with same blur amounts σ_{blur}. Before calculation of the standard deviation, the 83 metric scores are normalized to the interval $[0, 1]$, by dividing each score by the maximum score for each σ_{blur}. The results are shown in Table II, where the proposed PSI exhibits much lower deviations for different images at the same blur level, *i.e.*, the average performance gains over S_3 and FISH are 31.70% and 31.93%, respectively.

TABLE II
STANDARD DEVIATION OF THE SHARPNESS METRICS FOR BLURRED VERSIONS OF THE ORIGINAL IMAGES FROM LIVE, TID AND CSIQ. LOW DEVIATIONS EXPRESS INVARIANCE TO THE IMAGE CONTENT.

σ_{blur}	JNB	CPBD	S_3	FISH	PSI
0.25	0.1523	0.1017	0.1067	0.1215	**0.0638**
0.50	0.1714	0.1359	0.1190	0.1303	**0.0890**
0.75	0.1627	0.1617	0.1456	0.1264	**0.0925**
1.00	0.1556	0.1690	0.1312	0.1238	**0.0924**
1.25	0.1474	0.1834	0.1273	0.1296	**0.0930**

V. CONCLUSION

An effective method for the assessment of perceived sharpness has been presented. This Perceptual Sharpness Index (PSI) uses extracted edge gradient features by utilizing characteristics of the human visual system. The evaluation on public quality assessment databases shows that the PSI correlates highly with human subjective judgments and performs well for a wide range of blurriness, from slightly to heavily blurred images, with a large degree of invariance for different image contents. We have demonstrated that the PSI is computationally efficient making it applicable for many practical imaging problems, such as quality assessment, blind deconvolution, automatic image enhancement, and super-resolution.

REFERENCES

[1] D. Shaked and I. Tastl, "Sharpness measure: towards automatic image enhancement," in *IEEE International Conference on Image Processing*, 2005.

[2] G. Blanchet and L. Moisan, "An explicit sharpness index related to global phase coherence," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2012.

[3] S. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 8, pp. 824 –831, 1994.

[4] J. Sun, Z. Xu, and H.-Y. Shum, "Image super-resolution using gradient profile prior," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

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

[6] C. Vu, T. Phan, and D. Chandler, "S3: A spectral and spatial measure of local perceived sharpness in natural images," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 934 –945, 2012.

[7] N. D. Narvekar and L. J. Karam, "A no-reference image blur metric based on the cumulative probability of blur detection (CPBD)," *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2678 –2683, 2011.

[8] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: application to JPEG2000," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 163 – 172, 2004.

[9] C. Li, W. Yuan, A. Bovik, and X. Wu, "No-reference blur index using blur comparisons," *IET Electronics Letters*, vol. 47, no. 17, 2011.

[10] P. Vu and D. Chandler, "A fast wavelet-based algorithm for global and local image sharpness estimation," *IEEE Signal Processing Letters*, vol. 19, no. 7, pp. 423 –426, 2012.

[11] R. Hassen, Z. Wang, and M. Salama, "No-reference image sharpness assessment based on local phase coherence measurement," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2010.

[12] D. H. Hubel and T. N. Wiesel, "Early exploration of the visual cortex," *Neuron*, vol. 20, no. 3, pp. 401 – 412, 1998.

[13] A. Moorthy and A. Bovik, "Visual importance pooling for image quality assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 193 –201, 2009.

[14] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the dct domain," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339 –3352, 2012.

[15] P. Le Callet and F. Autrusseau, "Subjective quality assessment IRCCyN/IVC database," 2005, http://www.irccyn.ec-nantes.fr/ivcdb/.

[16] H. Sheikh, Z. Wang, L. Cormack, and A. Bovik, "LIVE image quality assessment database, release 2."

[17] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008 - a database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, vol. 10, pp. 30–45, 2009.

[18] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, 2010.

[19] A. Moorthy and A. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350 –3364, 2011.

[20] "Final report from the video quality experts group on the validation of objective models of video quality assessment." VQEG, 2000.