2021 IEEE International Conference on Image Processing (ICIP) | 978-1-6654-4115-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICIP42928.2021.9506498

AN EFFECTIVE SHARPNESS ASSESSMENT METHOD FOR SHALLOW DEPTH-OF-FIELD IMAGES

Zhixiang Duan[1], *Guangxin Li*[1,*], *Guoliang Fan*[2]

[1]School of Computer Science and Technology, Xidian University, Xi'an, China
[2]School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, USA
E-mail: 492734299@qq.com, lgx@xidian.edu.cn*, guoliang.fan@okstate.edu

ABSTRACT

No-reference (NR) image sharpness assessment is an important issue for image quality assessment and algorithm performance evaluation. Many objective NR sharpness assessment metrics have been proposed which are often intended to be strongly associated with the human visual system (HVS). However, recent studies show that common sharpness assessment indicators may misjudge the degree of blurring for images with shallow depth of field that are often used to highlight the main subject in the view. This paper proposes an efficient no-reference objective image sharpness assessment metric based on the product of bidirectional pixel intensity differences that is computed block-by-block (PBDB). This paper contributes the following: (1) the sharpness of shallow depth-of-field images can be accurately evaluated with the proposed algorithm when traditional methods do not work well. (2) Experimental results on three public datasets demonstrate competitiveness and effectiveness of the proposed algorithm when compared with several state-of-the-art methods.

Index Terms— Shallow Depth-of-Field Images, Image Sharpness Assessment, Image Block, Image Quality Assessment

1. INTRODUCTION

The sharpness of an image determines the amount of detail that an imaging system can present, and it is one of the most significant factors affecting the quality of an image. Subjective image quality metrics are considered the most reliable results in image assessment. Although subjective quality assessment methods can provide assessment results in line with human vision, subjective quality assessment methods still have high evaluation costs, low efficiency and other issues [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]. Therefore, based on the subjective nature of visual quality perception, to develop an accurate blind image quality assessment method becomes a challenging research topic. In recent years, researchers have proposed related artificial intelligence algorithms, but they still face the problem of high data collection and calculation costs [9, 10].

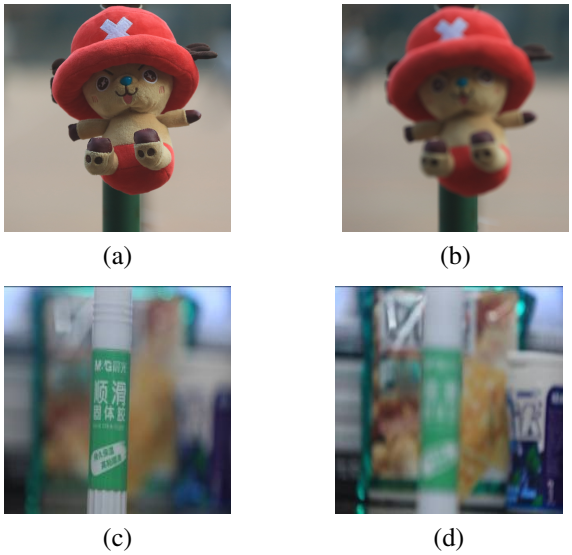

Fig. 1. (a) and (c) are pictures with shallow depth of field to highlight the subject. (b) and (d) are overall blurred pictures (the blur degree of the distant view is not as good as that in picture(a) and (c)).

A telephoto lens or a wide aperture lens can make the subject in a cluttered environment stand out. Photographers often push this effect further by using a very shallow depth of field while shooting, which blurs out everything unimportant and highlights the subject extremely well. The above-mentioned different types of lenses can get their shallow depth-of-field images respectively. The shallow depth-of-field images often exist in art, science, and daily life, so their sharpness assessment cannot be neglected.

Classic Image Quality Assessment (IQA) databases, such as LIVE [13] and TID2013 [14], only cover artificial and common distortion images, but do not include natural shallow depth-of-field images. Fig. 1.illustrates two types of photos for the same subject, in which picture (a) and (c) are two depth-of-field pictures taken by photographers to high-

*This research is supported in part by the National Key Research and Development Project of China under Grant 2016YFB1102500.

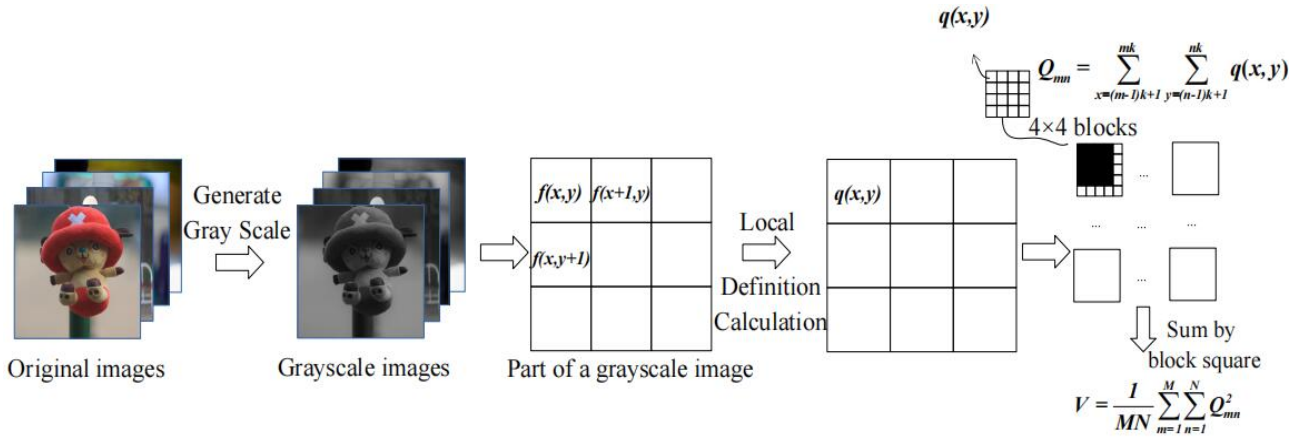

Fig. 2. Algorithm framework proposed

light the subject, and picture (b) and (d) are two global blur pictures (the blur degree of the distant view is not as good as that in picture(a) and (c)). Most people will think that picture (a) and (c) are clearer. However, the classic sharpness assessment algorithms may get the opposite result of HVS.

Scientific research shows that the highest visual acuity in the human visual system is limited by the size of the foveal area [2]. The foveal region is composed of an image block of approximately 64×64, which covers a view angle of about $\alpha = 2$, and is covered with cone cells for detecting brightness and color [2, 15, 16]. In this area, smooth blocks will be excluded because they do not cause blur. If the images with blurred background and clear foreground are considered, the clear foreground such as human faces attract people's visual attention, and then the correlation between the classic no-referenced sharpness indicators and the subjective scores will decrease [1, 2, 15, 16, 17].

To this end, this paper proposes a new no-referenced objective image sharpness assessment algorithm based on the square of block energy difference product, which can assess the sharpness of shallow depth-of-field images. Then, a dataset with shallow depth-of-field images is built for the sharpness assessment, and two comparative experiments are carried out for the assessment. And finally the paper presents experiment result from the proposed algorithm and other algorithms, which verifies the proposed algorithm's accuracy and universality.

2. PROPOSED METRIC

A new metric is proposed in this section to assess the sharpness of the shallow depth-of-field images. Studies show that the large changes in pixel intensity can indicate sharpness more than the small changes [5]. The difference product of gray values is used to amplify this difference and obtain excellent results. Fig. 2 shows the proposed algorithm framework. The specific implementation steps are as follows:

2.1. Grayscale block processing

Given a color image, we first obtain the gray scale matrix of the image for its assessment.

2.2. Local difference product matrix

For each pixel $f(x, y)$, calculate the gray value difference between the pixel and its adjacent pixel in x-axis, so does it in y-axis. The absolute product of the two value differences for each pixel on x-axis and y-axis represents the local sharpness index $q(x, y)$ of the pixel, and the formula is:

$$q(x, y) = |[f(x, y) - f(x + 1, y)] \times [f(x, y) - f(x, y + 1)]| \quad (1)$$

For further discussion, the following situations may be encountered when the difference product on the x-axis and y-axis is used:

a. Small value $*$ Small value $=$ Small value.
b. Large value $*$ Large value $=$ Large value.
c. Small value $*$ Large value $=$ Moderate value.

From the above three situations, the product will have a wider range. The absolute value of the gray value difference between the pixel and its adjacent pixel is in the range of 0–255, thus the value range of $q(x, y)$ 0–65025. We choose small values such as 1 and large values 255 as examples. In situation a, the product of the two small values is 1, which is still a very small value. In situation b, the product of the two large values is a very large value 65025. And in situation c, the product of a large value and a small value is a moderate value 255. Compared with using one gray value difference alone or adding two gray value differences, multiplication enlarges the gap between a larger difference value and a smaller one. At the same time, it enriches the contribution of a larger difference value to $q(x, y)$. Since the human visual system is more sensitive to the areas with large changes, the product of $q(x, y)$ is more suitable for it.

2.3. Sum of q within a block

The grayscale matrix is then divided into multiple $k \times k$ pixel matrix blocks, the default value is 4, and the total number of matrix blocks is recorded as $M \times N$. Because the image statistical characteristics are spatially non-stationary, and the image distortion is also spatially variable, the recommended block size is 3×3 and not greater than 6×6. The definition index of each pixel matrix block is numbered Q_{mn}. The sum

Fig. 3. Part of the experimental pictures shot by telephoto lens

of the local definition index q of each pixel in a certain pixel matrix block is calculated and recorded as the definition index Q_{mn} of the matrix block. The formula for calculating the definition index Q_{mn} of the pixel matrix block is:

$$Q_{mn} = \sum_{x=(m-1)k+1}^{mk} \sum_{y=(n-1)k+1}^{nk} q(x,y) \qquad (2)$$

where $m = 1, \ldots, M$ and $n = 1, \ldots, N$ for an image of size $Mk \times Nk$. And $q(x, y)$ represents the local definition of the pixel in the x-axis and y-axis of the current image.

2.4. Calculate the sum of squares of each block Q_{mn}

The formula for calculating the square sum of the definition index Qm for all matrix blocks is:

$$V = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} Q_{mn}^2 \qquad (3)$$

Among them, MN is the total number of matrix blocks, Vrepresenting the sharpness score of the entire picture.

For shallow depth of field images, which are partially clear and used to highlight the subject, the clearer regions contribute more sensibility than the smooth regions to HVS [15, 16]. The square of the block definition Q_{mn} index further amplifies the degree of local change on the basis of section 2.2. After the square Q_{mn} of the larger change, the advantage will be more prominent, while the flatter block will only have a smaller contribution.

3. EXPERIMENTAL RESULTS

3.1. Experiment Plan

In this part, performance results for our metric are presented. First, two data sets are selected for testing: for experiment A, 15 sets of shallow depth-of-field pictures taken with a telephoto lens are tested used to show how the sharpness of shallow depth-of-field images can be accurately evaluated with the proposed algorithm and how other algorithms fail. For experiment B, various blur types of image subsets are used from three large-size LIVE [13], TID2008 [18] and TID2013 [14] databases as testing beds to show how the proposed algorithm outperforms classic algorithms on the three public datasets with respect to universality and accuracy.

Second, such eight classic no-reference sharpness assessment methods are chosen for comparison as CPBD [3], MLV [5], Maxpol1, Maxpol2 [11], CurveletQA [7], BRISQUE [4], NIQE [12] and SSEQ [8].

Third, we refer to the suggestion given by the video quality experts group (VQEG) [19], and select a nonlinear mapping of the prediction results x to the subjective scores using the 5-parameter mapping function [13] mentioned by HR Sheikh in 2006 for fitting. We calculate SROCC (Spearman rank-order correlation coefficient indicates the prediction monotonicity). And PLCC(Pearson correlation coefficient indicates the prediction accuracy) are calculated after the predicted quality scores are passed through a nonlinear logistic mapping function. The following mapping function is widely used:

$$\text{Metric}(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2 (x - \beta_3)}} \right) + \beta_4 x + \beta_5 \qquad (4)$$

where $(\beta i | i = 1, 2, 3, 4, 5)$ are regression parameters to be found.

Metric	Algorithm failure times
Proposed Metric	0
Laplacian	7
CPBD [3]	7
Variance	6
MLV [5]	1
CurveletQA [7]	4
BRISQUE [4]	8
NIQE [12]	2
SSEQ [8]	12
Maxpol1 [11]	1
Maxpol2 [11]	1

Table 1. The classic NR IQA algorithms evaluate the number of failures in 15 sets of experimental data

3.2. Experiment Details

3.2.1. Performance Results Using a collection of depth-of-field pictures taken with a telephoto lens

In experiment A, Canon EOS 5D Mark III, equipped with a 300 mm long focal lens, and fixed f / 5.6, ISO 1250 for outdoor and indoor real shooting data, is chosen. A dataset is created, and part of the experimental pictures are shown in Fig. 3. The experimental pictures is divided into 15 groups, and tripod and light shield equipment are used to ensure the consistency of the shooting distance and shooting environment. Two pictures are included in each group to show how other algorithms fail. One is a shallow depth-of-field picture with clear foreground and blurred background to highlight the subject with the best focus effect, the other a whole blurred in the defocus state. For example, in Fig. 1, (a) and (b) are the first experimental group, and (c) and (d) is the second experimental group. We select 25 volunteers who have nothing to do with the project in the research institute. The subjective test process, in which evaluation attribute is set to sharpness, strictly adheres to the Performance Methods in Stimulus-comparison

Blur Type	Metric / DB	Criterion	Proposed Metric	CPBD [3]	MLV [5]	Curvelet QA[7]	BRISQUE [4]	NIQE [12]	SSEQ [8]	Maxpol1 [11]	Maxpol2 [11]
Gaussian blur	Tid2008	PLCC	**0.9258**	0.9008	0.9224	0.9056	0.6804	0.9248	0.9114	0.9246	0.9166
		SROCC	0.9114	0.8975	0.9075	**0.9250**	0.8290	0.9180	0.9073	0.9130	0.9170
	Tid2013	PLCC	**0.9339**	0.9012	0.9338	0.8688	0.7155	0.8507	0.8841	0.9034	0.8887
		SROCC	**0.9423**	0.8918	0.9200	0.8930	0.8010	0.8660	0.8860	0.8930	0.9030
	Live	PLCC	**0.9529**	0.9383	0.9506	0.8987	0.7831	0.9392	0.9422	0.9357	0.9360
		SROCC	0.9579	**0.9590**	0.9390	0.9301	0.9312	0.9317	0.9514	0.9509	0.9065
JPEG compression	Tid2008	PLCC	**0.9514**	0.3146	0.9446	0.8293	0.9384	0.9412	0.8564	0.4846	0.6265
		SROCC	0.9206	0.2946	0.9122	0.8880	0.9130	**0.9210**	0.8530	0.2852	0.6630
	Tid2013	PLCC	0.9392	0.4885	**0.9395**	0.9256	0.9327	0.9325	0.855	0.5314	0.7033
		SROCC	0.9139	0.4994	0.9016	0.8790	0.9010	**0.9220**	0.839	0.3715	0.7110
	Live	PLCC	0.9199	0.0538	0.8916	0.8192	**0.9295**	0.9008	0.9144	0.4989	0.4645
		SROCC	0.8792	0.0493	0.8595	0.8059	**0.8863**	0.8812	0.8723	0.3070	0.2054
JPEG2000 compression	Tid2008	PLCC	**0.9307**	0.9206	0.9224	0.7726	0.7961	0.8276	0.8380	0.5989	0.6750
		SROCC	**0.8947**	0.8621	0.8751	0.7480	0.7860	0.8260	0.8290	0.4610	0.6160
	Tid2013	PLCC	**0.9571**	0.9435	0.9500	0.8361	0.7849	0.8617	0.8931	0.6387	0.6958
		SROCC	**0.9353**	0.8937	0.9132	0.8000	0.8470	0.8750	0.8890	0.4860	0.6290
	Live	PLCC	0.9005	0.9024	0.8975	0.8243	**0.9155**	0.9152	0.8941	0.5290	0.5737
		SROCC	0.8872	0.8920	0.8824	0.8235	0.8902	**0.9022**	0.8881	0.4304	0.6128
Denoising	Tid2008	PLCC	**0.9530**	0.8443	0.9401	0.9095	0.7750	0.8219	0.9482	0.6658	0.9119
		SROCC	**0.9397**	0.8660	0.8881	0.886	0.7350	0.7850	0.8970	0.6050	0.8440
	Tid2013	PLCC	**0.9510**	0.8654	0.9443	0.9049	0.7565	0.7925	0.8996	0.4779	0.8682
		SROCC	**0.9299**	0.8724	0.9219	0.8240	0.7610	0.7660	0.8620	0.3210	0.8220
Chromatic aberrations	Tid2013	PLCC	**0.9731**	0.8857	0.9717	0.7867	0.7611	0.8344	0.7768	0.6984	0.8605
		SROCC	**0.9066**	0.8130	0.8688	0.7720	0.7450	0.6550	0.8010	0.5640	0.7190
Sparse sampling and reconstruction	Tid2013	PLCC	0.9527	0.8715	**0.9568**	0.8260	0.8055	0.8307	0.8769	0.7822	0.8631
		SROCC	**0.9532**	0.8414	0.9330	0.8410	0.7200	0.8480	0.9160	0.6860	0.8110

Table 2. Performance of our method and other classic sharpness assessment metrics about 6 distortion types on three databases

Methods (SC) in ITU-R BT.500-14 [20], and the observer's task is to judge which one is the sharper from two pictures in each group. The sharper picture is marked as 1 and the less one is marked as 0. In this experiment, classic NR algorithms are chosen to evaluate each group of images. The number of groups opposite to the subjective assessment results are chosen in the assessment process of each algorithm. The maximum is 15, and the minimum is 0.

It can be seen from Table 1 that failures exist in the other 10 algorithms except the proposed. The SSEQ algorithm fails 12 times, BRISQUE 8 times. The traditional algorithms Laplacian and variance fail in 7 groups and 6 groups respectively. CPBD fails 7 times, CurveletQA fails 4 groups. The rest three algorithms of NIQE, MLV, and Maxpol perform well, but they all have failures, and the failures do not occur in the same experimental group.

3.2.2. Performance Results for LIVE, TID2008 and TID2013 databases

In experiment B, six blur types of images are chosen, among which Gaussian blur, JPEG compression, and JPEG2000 compression all appear in LIVE, TID2008 and TID2013, Denoising in TID2008 and TID2013, Chromatic aberrations and Sparse sampling and reconstruction in TID2013. The 6 types of blur selected are highly correlated with the assessment of clarity. It is worth noting that, in order to ensure the objectivity and fairness of the assessment, the mean opinion scores (MOS) are normalized in those three databases. The MOS of a picture is divided by the average of the MOS of all pictures in the corresponding group, which result in the normalized MOS of the picture to participate in the fitting.

Table 2 shows, the universality of our algorithm and the improvement of prediction reliability are obvious for PLCC or SROCC. Among them, the proposed algorithm for Gaussian blur is ranked first in TID2013, and the top two in another database. The gap between the proposed algorithm and the first algorithm in JPEG compression is within 0.01. And in JPEG2000 compression, the proposed algorithm wins the first place in TID2008 and TID2013. In Denoising and Chromatic aberrations, the proposed algorithm ranks the first place. In Sparse sampling and reconstruction, the proposed algorithm and MLV are significantly better than other algorithms, and the proposed algorithm's SROCC is 2 percentage higher than that of MLV, which shows the better monotonicity of the proposed algorithm. The results show that the proposed algorithm can effectively improve the accuracy of sharpness assessment when evaluating the sharpness of different blur pictures.

4. CONCLUSION

In this paper, we propose an effective algorithm, based on the product of bidirectional pixel intensity differences that is computed block-by-block, which is inspired by the perception of HVS for sharpness assessment. The newly proposed algorithm obtains more accurate identification results when evaluating the shallow depth-of-field images. And the newly proposed algorithm shows competitive results over the traditional ones on three public image datasets as well.

5. REFERENCES

[1] Nabil G Sadaka, Lina J Karam, Rony Ferzli, and Glen P Abousleman, "A no-reference perceptual image sharpness metric based on saliency-weighted foveal pooling," in *2008 15th IEEE International Conference on Image Processing*. IEEE, 2008, pp. 369–372.

[2] Rony Ferzli and Lina J 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.

[3] Niranjan D Narvekar and Lina 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.

[4] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on image processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[5] Khosro Bahrami and Alex C Kot, "A fast approach for no-reference image sharpness assessment based on maximum local variation," *IEEE Signal Processing Letters*, vol. 21, no. 6, pp. 751–755, 2014.

[6] Yaqian Xu, Wenqing Zheng, Jingchen Qi, and Qi Li, "Blind image blur assessment based on markov-constrained fcm and blur entropy," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 4519–4523.

[7] Lixiong Liu, Hongping Dong, Hua Huang, and Alan C Bovik, "No-reference image quality assessment in curvelet domain," *Signal Processing: Image Communication*, vol. 29, no. 4, pp. 494–505, 2014.

[8] Lixiong Liu, Bao Liu, Hua Huang, and Alan Conrad Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Processing: Image Communication*, vol. 29, no. 8, pp. 856–863, 2014.

[9] Yabin Zhang, Haiqiang Wang, Fengfeng Tan, Wenjun Chen, and Zurong Wu, "No-reference image sharpness assessment based on rank learning," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 2359–2363.

[10] Yicheng Su and Jari Korhonen, "Blind natural image quality prediction using convolutional neural networks and weighted spatial pooling," in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020, pp. 191–195.

[11] Mahdi S Hosseini, Yueyang Zhang, and Konstantinos N Plataniotis, "Encoding visual sensitivity by maxpol convolution filters for image sharpness assessment," *IEEE Transactions on Image Processing*, vol. 28, no. 9, pp. 4510–4525, 2019.

[12] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal processing letters*, vol. 20, no. 3, pp. 209–212, 2012.

[13] Hamid R Sheikh, Muhammad F Sabir, and Alan C Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on image processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

[14] Nikolay Ponomarenko, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Lina Jin, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al., "Color image database tid2013: Peculiarities and preliminary results," in *european workshop on visual information processing (EUVIP)*. IEEE, 2013, pp. 106–111.

[15] Vivianne C Smith and Joel Pokorny, "Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm," *Vision research*, vol. 15, no. 2, pp. 161–171, 1975.

[16] R Siminoff, "Simulated bipolar cells in fovea of human retina," *Biological cybernetics*, vol. 66, no. 2, pp. 123–135, 1991.

[17] Ethan A Rossi and Austin Roorda, "The relationship between visual resolution and cone spacing in the human fovea," *Nature neuroscience*, vol. 13, no. 2, pp. 156–157, 2010.

[18] N Ponomarenko, V Lukin, K Egiazarian, Jaakko Astola, Marco Carli, and Federica Battisti, "Color image database for evaluation of image quality metrics," in *2008 IEEE 10th workshop on multimedia signal processing*. IEEE, 2008, pp. 403–408.

[19] Jochen Antkowiak, TDF Jamal Baina, France Vittorio Baroncini, Noel Chateau, France FranceTelecom, Antonio Claudio França Pessoa, FUB Stephanie Colonnese, Italy Laura Contin, Jorge Caviedes, and France Philips, "Final report from the video quality experts group on the validation of objective models of video quality assessment march 2000," 2000.

[20] ITU-R Recommendation BT.500-14, "Methodologies for the subjective assessment of the quality of television images," 2019.