



Krzysztof Okarma<sup>1,\*</sup>, Piotr Lech<sup>1</sup> and Vladimir V. Lukin<sup>2</sup>

- <sup>1</sup> Department of Signal Processing and Multimedia Engineering, West Pomeranian University of Technology in Szczecin, 70-313 Szczecin, Poland; piotr.lech@zut.edu.pl
- <sup>2</sup> Department of Information and Communication Technologies, National Aerospace University, 61070 Kharkov, Ukraine; lukin@ai.kharkov.com
- \* Correspondence: okarma@zut.edu.pl

**Abstract:** In the recent years, many objective image quality assessment methods have been proposed by different researchers, leading to a significant increase in their correlation with subjective quality evaluations. Although many recently proposed image quality assessment methods, particularly fullreference metrics, are in some cases highly correlated with the perception of individual distortions, there is still a need for their verification and adjustment for the case when images are affected by multiple distortions. Since one of the possible approaches is the application of combined metrics, their analysis and optimization are discussed in this paper. Two approaches to metrics' combination have been analyzed that are based on the weighted product and the proposed weighted sum with additional exponential weights. The validation of the proposed approach, carried out using four currently available image datasets, containing multiply distorted images together with the gathered subjective quality scores, indicates a meaningful increase of correlations of the optimized combined metrics with subjective opinions for all datasets.



# 1. Introduction

The increasing popularity and availability of relatively cheap cameras, as well as electronic mobile devices, equipped with visual sensors, undoubtedly causes a dynamic growth of applicability of image and video analysis in many tasks. Some obvious examples may be related to video surveillance, traffic monitoring, video inspection and diagnostics, video-based navigation of mobile robots, or even autonomous vehicles. Some other applications are related to non-destructive testing, data fusion from various sensors, and many others, also related to modern Industry 4.0 solutions. Another factor, influencing the growing popularity of image analysis, is the development of some freeware libraries, such as OpenCV, that makes it possible to perform many tasks in real-time, especially with hardware support provided by modern Graphics Processing Units (GPU).

Nevertheless, machine and computer vision algorithms typically utilize natural images, which may be subject to various distortions, occurring not only during their acquisition but also caused by, e.g., lossy compression or the presence of transmission errors. This situation is typical for modern electronic devices, such as cameras, phones, and some other gadgets where image data are subject to several nonlinear transformations before recording. In such a case, the ability to detect such distortions and assess the overall image quality is an important challenge given the reliability of the results obtained from their analysis.

In the recent several years, many objective image quality assessment (IQA) metrics have been proposed, which may be divided into three major groups: full-reference (FR), which require the knowledge of the original "pristine" image without any distortions, no-reference (NR) methods, also known as "blind" metrics and less popular reducedreference (RR) approaches, which assume a partial knowledge of the original (reference)



Citation: Okarma, K.; Lech, P.; Lukin, V. Combined Full-Reference Image Quality Metrics for Objective Assessment of Multiply Distorted Images. *Electronics* 2021, 10, 2256. https://doi.org/10.3390/ electronics10182256

Academic Editors: Luca Mesin and Byung Cheol Song

Received: 23 July 2021 Accepted: 10 September 2021 Published: 14 September 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). image. Although NR methods are the most desirable, their universality and correlation with subjective opinions of the human observers, provided as Mean Opinion Scores (MOS) or Differential MOS (DMOS) values in IQA databases, are typically significantly lower in comparison to FR methods. The more detailed analysis of many metrics and their comparisons for various widely accepted datasets containing reference and distorted images together with subjective quality scores may be found in some recent survey papers [1–4].

There are numerous attempts to improve the correlation between FR metrics and MOS (or DMOS). One way to do this is to design so-called combined metrics [5–8] that jointly employ several metrics (that we call elementary) in one or another way. In practice, one needs easily computable metrics and a simple way of combining them, similarly as for the 3D printed surfaces [9] or remote sensing images [10]. Because of this, the goal of this paper is to put forward a family of combined metrics that can be optimized with application to assessing the quality of images with multiple distortions. To the best of our knowledge, such optimization has not been yet carried out for available databases containing only images with multiple distortions. Previously developed combined metrics [5,6,8,11,12] concern only the singly distorted images.

The most commonly appearing types of distortions that an ideal IQA metric should be sensitive to concern blurring artifacts, various types of noise, and lossy compression artifacts. Although in some IQA datasets containing singly distorted images more than 20 types may be distinguished, e.g., 24 types in the TID2013 dataset [13] including colorrelated distortions, their combinations provided in the multiply-distorted IQA datasets are limited to a few kinds of them. Typically, they are the combinations of blur, noise, JPEG/JPEG 2000 artifacts, and contrast change. These five common types of distortions have been used, e.g., in the MDID database [14] discussed in Section 3.

Considering the interference of individual distortions and their influence on the perceived image quality, the usefulness of some metrics designed for singly distorted images for the development of the combined metrics highly correlated with subjective quality assessment of multiply distorted images is not obvious and should be verified experimentally.

The rest of the paper is organized as follows: Section 2 contains the overview of some elementary metrics, typically applied for the quality assessment of singly-distorted images, whereas four publicly available multiply-distorted image datasets used in experiments are presented in Section 3. Section 4 is related to the description of the idea of combined metrics and the proposed approach with experimental results discussed in Section 5. Section 6 concludes the paper.

# 2. Overview of Some Elementary Metrics

The performance of a combined metric depends on the following elements:

- The number of the combined elementary metrics;
- Which metrics are combined;
- How the metrics are combined;
- What images are used in testing.

Hence, we start by recalling modern elementary metrics.

Development of modern visual quality metrics, replacing the "classical" pixel-based approaches such as Mean Square Error (MSE) or Peak Signal-to-Noise Ratio (PSNR), started in fact in 2002 with the idea of the Universal Image Quality Index (UQI) [15], followed by its improvement widely known as the Structural SIMilarity (SSIM) [16], implemented also in the multi-scale version (MS-SSIM) [17].

The general formula describing the idea of the SSIM, sensitive to three main types of distortions, i.e., luminance, contrast and structural distortions, may be expressed as

$$SSIM = l(x,y) \cdot c(x,y) \cdot s(x,y) = \frac{2\bar{x}\bar{y} + C_1}{\bar{x}^2 + \bar{y}^2 + C_1} \cdot \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \cdot \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} , \qquad (1)$$

where the default values of the stabilizing constants (preventing the instability of results for dark and flat image areas) for 8-bit grayscale images are:  $C_1 = (0.01 \times 255)^2$ ,  $C_2 = (0.03 \times 255)^2$  and  $C_3 = C_2/2$ . The above computations are performed using the sliding window approach and the final metric is the average of the local similarities.

This approach was the basis also for some other similarity-based metrics leading to a further increase of the correlations between the objective quality scores and subjective MOS or DMOS values provided in various IQA datasets (typically containing only singlydistorted images). Some such examples, used also in this paper, are: information content weighted SSIM (IW-SSIM) and IW-PSNR [18], Complex Wavelet SSIM (CW-SSIM) [19], Feature SIMilarity (FSIM) [20], Quality Index based on Local Variance (QILV) [21] as well as a color version of SSIM (CSSIM), SSIM4 and its color version CSSIM4 [22], belonging to the group of SSIM-based metrics with additional predictability of image blocks.

A good illustration of the exemplary modifications of the SSIM might be the QILV metric [21] expressed as

$$QILV = \frac{2\mu_{V_A}\mu_{V_B}}{\mu_{V_A}^2 + \mu_{V_B}^2} \cdot \frac{2\sigma_{V_A}\sigma_{V_B}}{\sigma_{V_A}^2 + \sigma_{V_B}^2} \cdot \frac{\sigma_{V_A}v_B}{\sigma_{V_A}\sigma_{V_B}} , \qquad (2)$$

where  $\sigma_{V_A V_B}$  denotes the covariance between the variances of two images ( $V_A$  and  $V_B$ , respectively),  $\sigma_{V_A}$  and  $\sigma_{V_B}$  are the global standard deviations of the local variance with  $\mu_{V_A}$  and  $\mu_{V_B}$  being the mean values of the local variance.

Another example may be FSIM [20] based on the local similarity defined as

$$S_L(\mathbf{x}) = \left(\frac{2 \cdot PC_A(\mathbf{x}) \cdot PC_B(\mathbf{x}) + T_1}{PC_A^2(\mathbf{x}) + PC_B^2(\mathbf{x}) + T_1}\right)^{\alpha} \cdot \left(\frac{2 \cdot GM_A(\mathbf{x}) \cdot GM_B(\mathbf{x}) + T_2}{GM_A^2(\mathbf{x}) + GM_B^2(\mathbf{x}) + T_2}\right)^{\beta},$$
(3)

where  $T_1$  and  $T_2$  are the stability constants preventing the division by zero and **x** is the sliding window position. The two main components are the phase congruency (PC) being a significance measure of a local structure) and gradient magnitude (GM) as a complementary feature extracted using the Scharr edge filter. The final metric should be calculated according to the formula

$$FSIM = \frac{\sum_{\mathbf{x} \in A} S_L(\mathbf{x}) \cdot PC_m(\mathbf{x})}{\sum_{\mathbf{x} \in A} PC_m(\mathbf{x})}, \qquad (4)$$

where  $PC_m(\mathbf{x}) = max(PC_A(\mathbf{x}), PC_B(\mathbf{x}))$  and  $\mathbf{x}$  denotes each position of the local window on the image plane *A* (or *B*).

Another approach, originating from information theory, assumes the use of natural scene statistics (NSS) combined with a measurement of the mutual information between the subbands in the wavelet domain, proposed by Sheikh and Bovik as Visual Information Fidelity (VIF) metric [23]. Its simplified multi-scale pixel domain version (VIFp) requires fewer computations, although it does not allow the orientation analysis. Both methods are based on the earlier idea of Information Fidelity Criterion (IFC) [24]. A lower computational complexity metric, known as DCT Subbands Similarity (DSS) [25] utilizes the fact that statistics of DCT coefficients change with the degree and type of image distortion. Another motivation for its authors has been the popularity of the 2D DCT as many image and video coding techniques are based on block-based DCT transforms, particularly originating from JPEG and MPEG standards.

A combination of steerable pyramid wavelet transform and SSIM, known as IQM2, was proposed by Dumic et al. [26], where the kernel with two orientations was applied to achieve the best performance preserving low computational demands.

A different approach to the perceptual IQA was proposed by Wu et al. [27], utilizing the internal generative mechanism (IGM) adopting a Bayesian prediction model and decomposing the image into predicted and disorderly portions. It was assumed that the first part may be assessed using the SSIM-like methods, whereas the degradation on disorderly uncertainty may be predicted using the PSNR. Both parts should be further nonlinearly combined to acquire the final quality score.

Chang et al. [28] proposed the method based on the independent feature similarity (IFS) simulating the properties of the Human Visual System (HVS), particularly useful for the quality prediction of images with color distortions. Due to the possible use of the partial information from the reference image (based on the use of Independent Component Analysis—ICA), this method can also be considered as an example of the RR approach. Another metric based on the HVS, known as Perceptual SIMilarity (PSIM) was proposed as a four-step method [29] and partially verified using two multiply distorted databases. It is based on the extraction of gradient magnitude maps for both compared images followed by calculations of their multi-scale similarities and measurement of chromatic channel degradations and final pooling.

Alternatively, authors of the Sparse Feature Fidelity (SFF) metric [30] assumed transformation of images into sparse representations in the primary visual cortex to detect the sparse features by the feature detector trained by the ICA algorithm using natural image samples. They used feature similarity and luminance correlation components to simulate jointly visual attention and visual threshold. The other metric based on sparse representations, known as UNIQUE [31], utilized an unsupervised learning approach. Interestingly, in the preprocessing step, a color space selection is performed (conversion into YCbCr model is suggested with replacement of the Cb chrominance by the green channel) followed by random patch sampling, forming the vector containing 64 elements for each of three channels, further normalization using a mean subtraction and a whitening operation. The additional extension by analyzing the learned weights was proposed as the MS-UNIQUE metric [32]. Both metrics were trained using randomly selected patches from the ImageNet database. Further extension of such a training-based approach, particularly using deep learning CNN approaches [33,34], is also possible; however, it still requires a relatively large amount of training data available mainly in the singly distorted IQA datasets.

An interesting metric, utilizing gradient similarity, chromaticity similarity, and deviation pooling, was proposed as the Mean Deviation Similarity Index (MDSI) [35], where the color distortions were measured using a joint similarity map of two chromatic channels. Another attempt to use the gradient similarity has been proposed by Xue et al. [36], known as Gradient Magnitude Similarity Deviation (GMSD).

Reisenhofer et al. [37] proposed the use of the Haar wavelet decomposition to develop another HVS-based perceptual similarity metric, known as HaarPSI. This metric is based on the use of six 2D Haar wavelet filters extracting the horizontal and vertical edges on different frequency scales and may be considered as a simplification of FSIM [20]. Another feature-based method, known as RVSIM [38], utilizes Riesz transform (similarly as earlier RFSIM [39]) together with visual contrast sensitivity, whereas the CVSSI metric [40] is based on the similarity of contrast and visual saliency (VS), forming the final score with the use of weighted standard deviations of the local contrast quality map and the global VS quality map.

Considering the topic of this paper, the above overview of elementary metrics is limited to the FR algorithms demonstrating a high prediction accuracy for the four considered multiply distorted IQA datasets, obtained without any nonlinear fitting functions (e.g., logistic or polynomial ones). Although a few metrics oriented for the quality assessment of multiply distorted images have been recently proposed, e.g., using gradient detection [41], in some cases, their codes are not publicly available or they belong to the group of "blind" methods, such as the method based on phase congruency [42]. Therefore, the results presented in this paper are focused on the combination of better-known elementary metrics with available codes, originally developed for singly distorted images.

In addition to the above-mentioned metrics, some of the IQA methods, which have led to an improved performance applied in the combined metrics, include: WSNR [43], PSNRHMA [44], VSNR [45], Visual Saliency-Induced Index (VSI) [46], Multiscale Contrast

Similarity Deviation (MCSD) [47], spectral residual similarity (SR-SIM) [48] and Wavelet Based Sharp Features (WASH) [49]. Some other recently proposed metrics used in experiments have been developed originally for the quality estimation of screen content images, such as SIQAD [50] and SCI\_GSS [51], as well as for the reduced-reference image quality assessment of contrast change (RIQMC) [52].

Since some of the methods presented above are designed for the direct use with color images only and the others require the use of grayscale ones, all the calculations for the latter ones have been made using MATLAB's *rgb2gray* conversion, according to the ITU-R BT.601-7 Recommendation, after rounding to three decimal places.

# 3. Multiply Distorted Image Quality Assessment Datasets

The development of new IQA datasets is a quite challenging and time-consuming task, especially assuming conducting perceptual experiments involving many observers for a relatively large number of distorted images. Hence, among many IQA datasets, only a few of them, such as, e.g., TID2013 [13], containing numerous images subject to several types of distortions, may be considered as widely accepted by the community. Unfortunately, most of the databases developed several years ago do not contain images with more than a single distortion applied simultaneously, and most of the metrics developed and verified using such datasets predict the quality of multiply distorted images with relatively low accuracy.

As stated by Chandler [2], one of the main challenges in the multiply distorted IQA is the fact that the developed metrics should consider not only the joint effects of distortions on the image but also the effects of distortions on each other. Hence, considering the practical usefulness of metrics that would be able to predict the visual quality of multiply distorted images with the possibly highest accuracy, some other datasets have been developed to fill this research gap.

The first of such datasets, provided by the Laboratory for Image and Video Engineering (LIVE) from Texas University at Austin, referred to as LIVEMD [53], contains two groups of doubly distorted images. The first group deals with a blur followed by JPEG lossy compression, whereas the second one contains blurred images due to defocusing corrupted further by a white noise to simulate sensor noise. Each group contains 225 images, however, some of them are in fact singly distorted, hence only the subset of 270 multiply distorted images has been used in experiments carried out in our paper.

Another dataset, known as MDID13 [54], contains 12 natural color reference images and 324 images corrupted simultaneously by distortions that may take place during the acquisition, compression, and transmission of images. Six standard definition reference images ( $768 \times 512$  pixels) originate from the Kodak database, whereas the other six high definition images ( $1280 \times 720$ ) are the same as in the LIVEMD dataset. The testing images contain the three-fold mixtures of blurring, JPEG compression, and noise, being complementary to the LIVEMD, where only two-fold artifacts are used. Subjective scores have been provided by 25 inexperienced observers using two viewing distances due to different image sizes and the single-stimulus (SS) method according to the ITU-R BT.500-12 Recommendation.

The third database used for the verification of the proposed approach is known simply as MDID [14]. It contains 20 reference images (cropped to 512 × 384 pixels without scaling) and 1600 distorted images. The images are corrupted by the combinations of five distortions, namely Gaussian noise (GN), Gaussian blur (GB), contrast change (CC), JPEG, and JPEG2000 lossy compression. Each distorted image has been obtained from the respective reference image applying random types and random levels of distortions. The MOS values have been provided by 192 subjects who participated in the subjective rating. Sample images from the MDID database affected by various combinations of distortions with different levels are presented in Figure 1 with the reference image marked by the red frame.



**Figure 1.** Sample images from the MDID database [14]: (a) "pristine" image no. 8; (b) distorted by Gaussian blur (GB), contrast change (CC), JPEG lossy compression, and Gaussian noise (GN); (c) distorted by CC, GB, and JPEG; (d) distorted by GB, JPEG2000 lossy compression, and GN; (e) distorted by GB, JPEG, and GN; (f) distorted by CC, GB, JPEG2000, and GN; (g) distorted by JPEG2000; (h) distorted by JPEG2000 and GN; (i) distorted by GB, CC, and JPEG2000.

The last dataset, developed in the Imaging and Vision Laboratory at the University of Milano-Bicocca, is known as IVL\_MD or MDIVL database [55]. It contains two groups of images: 400 images with noise and JPEG distortions, as well as 350 images with blur plus JPEG distortions, together with corresponding MOS values. The distorted images, subjectively evaluated by 12 observers using the SS method, have been obtained from 10 reference images that have the size of 886 × 591 pixels.

There are also other databases containing images with multiple distortions, e.g., LIVE in the Wild Image Quality Challenge database, containing widely diverse authentic image distortions [56]. However, this database does not offer reference images and, therefore, it does not allow calculating FR metrics that are needed in our case.

Comparing the four publicly available multiply distorted IQA databases, the most relevant one is undoubtedly the MDID database [14], not only because of the largest number of images and distortion types but also considering the numerous human observers involved in perceptual experiments. Therefore, the experimental results obtained for this dataset should be considered as the most important. On the other hand, due to the greater diversity of distortions and higher number of images, the expected correlation values are lower than for the other datasets.

To provide a comparison of the performance of the best elementary (individual) metrics for each of the above databases, the Pearson Linear Correlation Coefficients (PCC) between the raw objective scores (i.e., without any additional nonlinear fitting) and subjective MOS/DMOS values have been calculated, illustrating the prediction accuracy. Additionally, Spearman Rank Order Correlation Coefficients (SROCC) and Kendall Rank Order Correlation Coefficients (KROCC) have been calculated to illustrate the prediction monotonicity of each elementary metric.

The obtained performance for selected elementary metrics, including the best performing ones, is presented in Table 1, where the top three results for each dataset are marked with bold font. As can be easily noticed, various methods demonstrate the best performance for various datasets, also differing with prediction accuracy measured by PCC and prediction monotonicity indicated by rank order correlations. Although not all results obtained for elementary metrics have been provided in the paper, the values of over 50 of them have been calculated for four considered datasets. Additionally, the correlation results obtained for all databases weighted by the number of images in each of the considered datasets have been presented. Therefore, the weights (before normalization) are 270 for LIVEMD excluding the single distorted part of the database), 324 for MDID13, 1600 for MDID, and 750 for MDIVL, respectively. Hence, the most "universal" elementary metrics seem to be VIF, DSS, and IW-SSIM, providing the highest aggregated correlations, being a good starting point for the development of the combined metrics.

**Table 1.** Performance of some elementary metrics (expressed as Pearson, Spearman, and Kendall correlation coefficients) for the considered IQA databases with multiply distorted images together with the average performance weighted by the size of individual datasets. The top three results for each dataset are marked with bold font.

Matria	Detabase	Correlation Coefficie			ents		Correlation Coefficients		
Metric	Database	PCC	SROCC	KROCC	Metric	Database	PCC	SROCC	KROCC
	LIVEMD	0.5082	0.5111	0.3603		LIVEMD	0.7398	0.7377	0.5298
IW-PSNR	MDID13	0.7649	0.7816	0.5697	IW-SSIM	MDID13	0.8413	0.8551	0.6574
[18]	MDID	0.6859	0.6719	0.4846	[18]	MDID	0.8634	0.8911	0.7092
	MDIVL	0.8303	0.8178	0.6229		MDIVL	0.6955	0.8588	0.6708
	Weighted	0.6738	0.7064	0.5178		Weighted	0.8069	0.8648	0.6773
	LIVEMD	0.6954	0.6922	0.4803		LIVEMD	0.6664	0.6909	0.4850
FSIM	MDID13	0.5697	0.5818	0.3899	CSSIM4	MDID13	0.8147	0.8628	0.6665
[20]	MDID	0.8597	0.8873	0.7077	[22]	MDID	0.5672	0.6639	0.4793
	MDIVL	0.7123	0.8589	0.6701		MDIVL	0.6326	0.9084	0.7320
	Weighted	0.7743	0.8275	0.6415	-	Weighted	0.6202	0.7505	0.5648
	LIVEMD	0.7709	0.7588	0.5428		LIVEMD	0.7051	0.7142	0.5061
VIF	MDID13	0.8221	0.8447	0.6440	VIFp	MDID13	0.7361	0.7594	0.5561
[23]	MDID	0.8873	0.9306	0.7714	[23]	MDID	0.8184	0.8770	0.6978
	MDIVL	0.8568	0.8378	0.6471		MDIVL	0.8000	0.7711	0.5721
	Weighted	0.8617	0.8817	0.7048	-	Weighted	0.7943	0.8221	0.6326
	LIVEMD	0.7070	0.7439	0.5453		LIVEMD	0.5087	0.6247	0.4305
DSS	MDID13	0.7907	0.8078	0.5950	IQM2	MDID13	0.7668	0.7806	0.5838
[25]	MDID	0.8711	0.8658	0.6788	[26]	MDID	0.8463	0.8530	0.6652
	MDIVL	0.8276	0.8759	0.6910		MDIVL	0.8681	0.8764	0.6891
	Weighted	0.8361	0.8508	0.6604	-	Weighted	0.8121	0.8300	0.6408
	LIVEMD	0.5527	0.6633	0.4606		LIVEMD	0.6668	0.6729	0.4763
IGM	MDID13	0.8007	0.8239	0.6241	IFS	MDID13	0.7132	0.7325	0.5305
[27]	MDID	0.8271	0.8548	0.6678	[28]	MDID	0.9007	0.9070	0.7367
	MDIVL	0.7872	0.8637	0.6728		MDIVL	0.7032	0.8296	0.6388
	Weighted	0.7889	0.8361	0.6453	•	Weighted	0.8083	0.8466	0.6652
	LIVEMD	0.6883	0.6920	0.4800		LIVEMD	0.7059	0.6940	0.4842
PSIM	MDID13	0.8325	0.8618	0.6630	MDSI	MDID13	0.6725	0.7024	0.4951
[29]	MDID	0.8427	0.8733	0.6871	[35]	MDID	0.8249	0.8360	0.6519
	MDIVL	0.7111	0.8427	0.6463		MDIVL	0.8297	0.8376	0.6449
	Weighted	0.7939	0.8476	0.6550	-	Weighted	0.7985	0.8087	0.6175
	LIVEMD	0.6094	0.7155	0.5187		LIVEMD	0.7139	0.7064	0.4835
HaarPSI	MDID13	0.8385	0.8470	0.6425	RVSIM	MDID13	0.6957	0.7253	0.5196
[37]	MDID	0.8922	0.8879	0.7125	[38]	MDID	0.8831	0.8835	0.7086
	MDIVL	0.7936	0.8140	0.6212		MDIVL	0.8626	0.8517	0.6596
	Weighted	0.8352	0.8487	0.6637	•	Weighted	0.8417	0.8418	0.6547

Matula	D ( 1	<b>Correlation Coefficients</b>		Matula	Detabase	<b>Correlation Coefficients</b>			
wietite	Database	PCC	SROCC	KROCC	Wietric	Database	PCC	SROCC	KROCC
	LIVEMD	0.7059	0.7303	0.5266		LIVEMD	0.7205	0.7261	0.5197
CVSSI	MDID13	0.7903	0.8065	0.5959	SFF	MDID13	0.7887	0.8005	0.5931
[40]	MDID	0.8594	0.8638	0.6840	[30]	MDID	0.8047	0.8396	0.6599
	MDIVL	0.8098	0.8540	0.6659		MDIVL	0.7398	0.8535	0.6624
	Weighted	0.8239	0.8427	0.6552	-	Weighted	0.7787	0.8284	0.6403
	LIVEMD	0.7005	0.7417	0.5357		LIVEMD	0.7229	0.7241	0.5120
UNIQUE	MDID13	0.7004	0.8021	0.5983	MS- UNIOUE	MDID13	0.7274	0.8316	0.6312
[31]	MDID	0.7691	0.7944	0.5888	[32]	MDID	0.7245	0.7423	0.5407
	MDIVL	0.7678	0.7438	0.5498		MDIVL	0.7775	0.7550	0.5592
	Weighted	0.7549	0.7775	0.5751	-	Weighted	0.7382	0.7537	0.5528

Table 1. Cont.

#### 4. Combined Metrics and the Proposed Approach

Ideally, an FR metric has to provide a linear dependence between metric values and MOS. Less strictly, dependence between MOS and a metric should be monotonous (desirably, a larger metric value corresponds to a larger MOS). However, for many existing elementary metrics, these dependences are far from ideal. As examples, Figure 2 presents scatter plots of MOS vs. some elementary FR metrics for the considered databases (scatter plots in the left column). As one can see, the dependences can be nonlinear (as shown in the scatter plot of IQM2 vs. MOS), different metrics have different ranges of variation (many metrics vary in the limits from 0 to 1 but not all), some "outliers" (large displacements of some points with respect to the most of the others) might happen as well. These properties arise problems in aggregation of several elementary metrics into a combined one.

The idea of the combined metrics is motivated by the complementary properties of different elementary metrics, which may demonstrate a "sensitivity" to various kinds of distortions to varying degrees. Hence, it has been assumed that their nonlinear combination may replace the necessity of nonlinear fitting proposed by the Video Quality Experts Group (VQEG) to increase the linear correlation between the subjective and objective scores. Some initial attempts were made to combine the metrics for singly distorted images by the optimization of weighting exponents for the product of three metrics [5] using the TID2008 database, although during further experiments, one of the metrics was replaced by FSIM forming the Combined Image Similarity Index (CISI) [6], being the weighted product of MS-SSIM [17], VIF [18] and FSIM [20].

A multi-metric fusion based on the regression approach applied for some older elementary metrics was proposed in the paper [7] with the additional context-dependent version utilizing the machine learning approach to determine the context automatically. Nevertheless, the verification of results was made using the TID2008 dataset only.

Another approach to multi-metric fusion is based on the use of genetic algorithms for the combination of metrics [11], although modeled as their weighted sum instead of their product that may limit the possibility of avoiding the additional nonlinear fitting. Hence, a similar approach was also used for the weighted products of elementary metrics [12], leading to further improvements.

The use of neural networks for the combination of elementary IQA metrics was used in the paper [8], where a randomly selected half of the TID2013 dataset was used for training. This approach utilized six elementary metrics, leading to a significant increase of the SROCC chosen as the optimization criterion. Nevertheless, similarly as in the other cases, the combined metrics have been used only for the assessment of singly distorted images. Additionally, a potential application of deep learning methods would require the development of larger training datasets containing also the subjective quality scores for multiply distorted images. Therefore, a combination of existing metrics using a relatively simple model is expected to be a well-performing solution also for multiply distorted images.



**Figure 2.** Scatter plots for the "best" elementary metrics obtained for each considered dataset (**left column**) together with the plots generated for the combined CM7 (**middle column**) and CM<sup>+</sup>7 metrics (**right column**); from top to bottom for: LIVEMD, MDID13, MDID, and MDIVL databases. Subjective quality scores are expressed as MOS and DMOS whereas CM and CM<sup>+</sup> denote the objective combined metrics.

To provide a simple form of the combined metric which would not require the additional nonlinear regression, e.g., using the logistic function, the strategy based on the weighted product of elementary metrics has been initially chosen in this paper with PCC as the optimization criterion. Although, in some cases, prediction monotonicity may be more important than the prediction accuracy itself, we have verified experimentally that the optimization of weighting exponents using the PCC values as the criterion, provides also high SROCC values. During the experiments, it has appeared that the performances obtained in the opposite case are not always good enough. Another reason for the use of the PCC for raw scores without prior nonlinearity fitting was the flexibility of the proposed approach, making it possible to control all weights simultaneously in a single optimization procedure. Considering the various dynamic ranges of elementary metrics, as well as the DMOS and MOS values in each dataset, the use of the PCC does not require additional normalization of their values. Hence, the assumed formula of the combined metric may be expressed as:

$$CM = \prod_{i=1}^{N} Q_i^{w_i} , \qquad (5)$$

where *N* is the number of elementary metrics denoted as  $Q_i$ , and  $w_i$  are their exponential weights, obtained as the result of optimization conducted using MATLAB's *fminsearch* function.

Although the application of the assumed method of metrics' combination provides encouraging results, the selected fusion of metrics based on their weighted product does not always lead to fully satisfactory performance. Hence, a novel fusion model has been investigated based on the sum of the exponentially weighted metrics where each component of the sum has an additional weight. The proposed formula may be presented as:

$$CM^{+} = \sum_{i=1}^{N} (a_{i} \cdot Q_{i}^{w_{i}})$$
, (6)

where the additional weights  $a_i$  have been introduced to make the combined metric even more flexible and increase its correlation with subjective quality scores provided in stateof-the-art datasets for multiply distorted images.

# 5. Results of Optimization

Using the weights a in Equation (6), different ranges of metrics' variation are taken into account (i.e., specific normalization is performed). Using both a and w coefficients, the combined metric can be optimized, i.e., its better values of PCC and/or SROCC can be provided in comparison to elementary metrics used as inputs for the combined metric.

An initial verification of the usefulness of the proposed approach for the FR quality assessment of multiply distorted images has been made primarily for the metrics listed in Table 1 using the four considered datasets independently. All initially considered metrics providing the PCC values below the bottom limits assumed for all datasets have been excluded from initial experiments (i.e., at least one of the conditions should be fulfilled by each metric to be included in further experiments). The values of these limits for PCC are: 0.7 for LIVEMD, 0.8 for MDID13, 0.85 for MDID and 0.8 for MDIVL. The relatively low limit for the LIVEMD dataset is caused by removing the singly distorted images from the analysis leading to a decrease of the correlation values for this dataset. Nevertheless, in some cases, combinations of two or three "worse" metrics might provide better results in comparison to the combination of one of them with the best performing elementary metric. Therefore, in the second stage of experiments, all combinations of two and three metrics have been tested for all datasets. To limit the number of possible combinations reasonably, several "best" combinations have been chosen as the basis for further increase of the number of metrics.

The optimization of exponential parameters  $w_i$  for the combined metrics CM as well as the multipliers  $a_i$  and exponents  $w_i$  for the proposed CM<sup>+</sup> formula has been conducted using the derivative-free method without constraints based on the Nelder–Mead simplex method implemented in MATLAB's *fminsearch* function. Finally, all multipliers  $a_i$  in the proposed CM<sup>+</sup> formula have been normalized so that  $\sum a_i = 1$ .

As the "best" combinations of two, three and more metrics for individual databases differ from each other, they are presented in Table 2 separately for each dataset. Analyzing the obtained results, it can be noticed that a meaningful increase of the prediction accuracy has been achieved for all datasets even using the "best" combination of two or three elementary metrics using the weighted product of metrics denoted as CM. The use of more additional elementary metrics further improves the obtained results in terms of the PCC significantly and, in some cases, may lead to a slight decrease of the prediction monotonicity (lower values of SROCC and KROCC).

Database	Metrics	Cor	relation Coeffic	ients	Denotation
		PCC	SROCC	KROCC	2011000000
LIVEMD	IFC	0.7871	0.7891	0.5869	(elementary)
	IW-SSIM, CSSIM	0.8637	0.8669	0.6741	CM2 <sub>LIVEMD</sub>
	FSIM, IW-SSIM, SSIM4	0.8880	0.8853	0.7040	CM3 <sub>LIVEMD</sub>
	FSIM, IW-SSIM, SSIM4, GMSD	0.8967	0.8900	0.7097	CM4 <sub>LIVEMD</sub>
	FSIM, IW-SSIM, SSIM4, GMSD, CSSIM	0.9055	0.9037	0.7316	CM5 <sub>LIVEMD</sub>
	FSIM, IW-SSIM, SSIM4, GMSD, CSSIM, UNIQUE	0.9132	0.9107	0.7406	CM6 <sub>LIVEMD</sub>
	FSIM, IW-SSIM, SSIM4, GMSD, CSSIM, UNIQUE, CSSIM4	0.9171	0.9135	0.7435	CM7 <sub>LIVEMD</sub>
MDID13	IW-SSIM	0.8413	0.8551	0.6574	(elementary)
	VSNR, CSSIM4	0.8930	0.9007	0.7159	CM2 <sub>MDID13</sub>
	PSIM, VSNR, CSSIM4	0.9133	0.9171	0.7418	CM3 <sub>MDID13</sub>
	PSIM, VSNR, CSSIM4, WSNR	0.9193	0.9214	0.7506	CM4 <sub>MDID13</sub>
	PSIM, VSNR, CSSIM4, WSNR, RIQMC	0.9235	0.9261	0.7606	CM5 <sub>MDID13</sub>
	PSIM, VSNR, CSSIM4, WSNR, RIQMC, CVSSI	0.9280	0.9304	0.7649	CM6 <sub>MDID13</sub>
	PSIM, VSNR, CSSIM4, WSNR, RIQMC, SR-SIM, FSIM	0.9342	0.9370	0.7769	CM7 <sub>MDID13</sub>
MDID	IFS	0.9007	0.9070	0.7367	(elementary)
	IFC, MCSD	0.9456	0.9478	0.7999	CM2 <sub>MDID</sub>
	IFC, MCSD, UQI	0.9520	0.9545	0.8132	CM3 <sub>MDID</sub>
	IFC, MCSD, UQI, QILV	0.9542	0.9566	0.8173	CM4 <sub>MDID</sub>
	IFC, MCSD, UQI, QILV, MS-UNIQUE	0.9559	0.9586	0.8215	CM5 <sub>MDID</sub>
	IFC, MCSD, UQI, QILV, MS-UNIQUE, RVSIM	0.9579	0.9608	0.8259	CM6 <sub>MDID</sub>
	IFC, MCSD, UQI, QILV, MS-UNIQUE, RVSIM, IW-SSIM	0.9587	0.9606	0.8261	CM7 <sub>MDID</sub>
MDIVL	IQM2	0.8681	0.8764	0.6891	(elementary)
	SIQAD, CSSIM4	0.9400	0.9142	0.7431	CM2 <sub>MDIVL</sub>
	QILV, SR-SIM, CSSIM4	0.9474	0.9291	0.7659	CM3 <sub>MDIVL</sub>
	QILV, SR-SIM, CSSIM4, SIQAD	0.9502	0.9292	0.7675	CM4 <sub>MDIVL</sub>
	QILV, SR-SIM, CSSIM4, SIQAD, CW-SSIM	0.9537	0.9410	0.7866	CM5 <sub>MDIVL</sub>
	QILV, SR-SIM, CSSIM4, SIQAD, CW-SSIM, PSNRHMA	0.9553	0.9429	0.7901	CM6 <sub>MDIVL</sub>
	QILV, SR-SIM, CSSIM4, SIQAD, CW-SSIM, PSNRHMA, VSI	0.9560	0.9441	0.7923	CM7 <sub>MDIVL</sub>

**Table 2.** Performance of the "best" elementary and combined metrics CM expressed as Pearson, Spearman, and Kendall correlation coefficients for the considered IQA databases with multiply distorted images.

The results of the application of the proposed CM<sup>+</sup> metrics based on the normalized sum of the exponentially weighted elementary metrics are presented in Table 3, where higher correlations in comparison to respective CM metrics are marked by bold font. As may be noticed, the obtained performance of the proposed combined metrics is better for three datasets and slightly worse for the MDID database. An additional comparison of the linearity of the achieved correlation (without the necessity of any additional nonlinear mapping) is presented in the scatter plots shown in Figure 2.

However, it should be kept in mind that many elementary metrics have various properties and various dynamic ranges, hence, the trends shown in the various plots may be reversed to each other. For some of these metrics, smaller values indicate higher quality whereas the opposite is true for some other metrics. Since the maximum absolute value of the PCC has been considered as the objective function, the presentation of the scatter plots using the raw scores of these metrics may present both "negative" and "positive" trends. It is dependent on the obtained results of the optimization and the elementary metrics which have been used in the final combined metric. As in two datasets the DMOS values have been provided as the subjective scores, whereas the inventors of the other two datasets have used the MOS values, the original values—different for different datasets—have been used in the paper and are presented in all scatter plots included in the paper. The scale of all obtained results have not been normalized. It should also be noted that the high DMOS values typically represent poor quality whereas high MOS values indicate a high quality of images.

**Table 3.** Performance of the "best" elementary and combined metrics CM<sup>+</sup> expressed as Pearson, Spearman and Kendall correlation coefficients for the considered IQA databases with multiply distorted images. Higher correlations in comparison to respective CM metrics are marked by bold font.

Database	Metrics	Corre PCC	elation Coeff SROCC	icients KROCC	Denotation
LIVEMD	IFC	0.7871	0.7891	0.5869	(elementary)
	IW-PSNR, SCI_GSS	0.8512	0.8498	0.6536	$CM^{+}2_{LIVEMD}$
	FSIM, IW-SSIM, SSIM	0.8732	0.8720	0.6844	$CM^{+}3_{LIVEMD}$
	FSIM, IW-SSIM, SSIM, SSIM4	0.9075	0.9042	0.7359	$CM^+4_{LIVEMD}$
	FSIM, IW-SSIM, SSIM, SSIM4, UNIQUE	0.9118	0.9047	0.7390	$CM^+5_{LIVEMD}$
	FSIM, IW-SSIM, SSIM, SSIM4, UNIQUE, IQM2	0.9299	0.9231	0.7621	$CM^+6_{LIVEMD}$
	FSIM, IW-SSIM, SSIM, SSIM4, UNIQUE, IQM2, CVSSI	0.9357	0.9302	0.7738	$CM^{+}7_{LIVEMD}$
MDID13	IW-SSIM	0.8413	0.8551	0.6574	(elementary)
	VSNR, CSSIM4	0.9013	0.9053	0.7253	$CM^+2_{MDID13}$
	VSNR, PSIM, MS-UNIQUE	0.9228	0.9247	0.7577	$CM^+3_{MDID13}$
	VSNR, PSIM, MS-UNIQUE, WSNR	0.9272	0.9260	0.7636	$CM^+4_{MDID13}$
	VSNR, PSIM, MS-UNIQUE, WSNR, SIQAD	0.9329	0.9319	0.7727	$CM^+5_{MDID13}$
	VSNR, PSIM, MS-UNIQUE, WSNR, SIQAD, QILV	0.9372	0.9347	0.7742	$CM^+6_{MDID13}$
	VSNR, PSIM, MS-UNIQUE, WSNR, SIQAD, QILV, RFSIM	0.9422	0.9423	0.7901	CM <sup>+</sup> 7 <sub>MDID13</sub>
MDID	IFS	0.9007	0.9070	0.7367	(elementary)
	IFC, MCSD	0.9447	0.9459	0.7955	$CM^+2_{MDID}$
	IFC, IFS, WASH	0.9517	0.9513	0.8029	$CM^+3_{MDID}$
	IFC, IFS, WASH, VSI	0.9521	0.9534	0.8077	$CM^+4_{MDID}$
	IFC, IFS, WASH, VSI, SSIM	0.9552	0.9569	0.8154	$CM^+5_{MDID}$
	IFC, IFS, WASH, VSI, SSIM, IW-SSIM	0.9574	0.9581	0.8180	$CM^+6_{MDID}$
	IFC, IFS, WASH, VSI, SSIM, IW-SSIM, MS-UNIQUE	0.9581	0.9594	0.8205	CM <sup>+</sup> 7 <sub>MDID</sub>
MDIVL	IQM2	0.8681	0.8764	0.6891	(elementary)
	SIQAD, CSSIM4	0.9381	0.9098	0.7372	$CM^+2_{MDIVL}$
	DSS, QILV, SSIM4	0.9510	0.9485	0.7975	$CM^+3_{MDIVL}$
	DSS, QILV, SSIM4, IW-PSNR	0.9529	0.9500	0.8013	$CM^+4_{MDIVL}$
	DSS, QILV, SSIM4, IW-PSNR, CSSIM4	0.9586	0.9581	0.8169	$CM^+5_{MDIVL}$
	DSS, QILV, SSIM4, IW-PSNR, CSSIM4, SIQAD	0.9606	0.9575	0.8168	$CM^+6_{MDIVL}$
	DSS, QILV, SSIM4, IW-PSNR, CSSIM4, SIQAD, CW-SSIM	0.9625	0.9608	0.8249	$CM^+7_{MDIVL}$

As it may be observed, results of the CM7<sup>+</sup> metric obtained for the MDID2013 dataset vary noticeably less than for the three other databases. Nevertheless, highly linear relationships between the subjective and objective quality scores are achieved mainly for the proposed CM<sup>+</sup> metrics for all considered databases. Some differences in the dynamic ranges of the combined metrics, particularly using the CM formulas, result from the use of various types of metrics and different weights obtained after the optimization procedure.

An additional comparison of the performance of the proposed approach has been made using some other combined metrics, previously developed for singly distorted images, applied for the datasets containing only multiply distorted images. The obtained experimental results for three such datasets (MDID2013, MDID, and MDIVL) are presented in Table 4. Since four Regression-based Similarity (rSIM) metrics [11] have been actually designed as the weighted sum of individual metrics, the additional nonlinear regression with the use of the logistic function has been applied using the coefficients provided in [11]. As one can see, our approach provides sufficiently better results than the approaches proposed in [11,12].

Since the metrics used in "best" combinations for various datasets differ, an additional cross-database validation has been conducted applying the combined metrics optimized for a single database for the assessment of images from the other three datasets. The obtained validation results are presented in Table 5, where the better performance results than obtained for the best elementary metrics for each dataset are marked with bold font. As it may be observed, the application of some of the combined metrics obtained for the MDIVL dataset does not lead to satisfactory results for the others.

Database	Motrico	Correlation Coefficients				
	wietrics	PCC	SROCC	KROCC		
MDID13	IW-SSIM	0.8413	0.8551	0.6574		
	CISI [6]	0.6882	0.6974	0.4894		
	rSIM1 [11]	0.7416	0.7487	0.5454		
	rSIM2 [11]	0.7438	0.7511	0.5529		
	rSIM3 [11]	0.7469	0.7519	0.5471		
	rSIM4 [11]	0.7464	0.7516	0.5476		
	ESIM1 [12]	0.5807	0.5858	0.4030		
	ESIM2 [12]	0.6666	0.6828	0.4794		
	ESIM3 [12]	0.7034	0.7316	0.5250		
	ESIM4 [12]	0.5773	0.5915	0.4015		
	CM <sup>+</sup> 7 (best proposed)	0.9422	0.9423	0.7901		
MDID	IFS	0.9007	0.9070	0.7367		
	CISI [6]	0.9045	0.9116	0.7427		
	rSIM1 [11]	0.7443	0.7266	0.5344		
	rSIM2 [11]	0.7429	0.7227	0.5320		
	rSIM3 [11]	0.7453	0.7259	0.5342		
	rSIM4 [11]	0.7442	0.7251	0.5334		
	ESIM1 [12]	0.8704	0.8641	0.6805		
	ESIM2 [12]	0.8780	0.8965	0.7247		
	ESIM3 [12]	0.8977	0.9114	0.7448		
	ESIM4 [12]	0.8752	0.8871	0.7089		
	CM7 (best proposed)	0.9587	0.9606	0.8261		
MDIVL	IQM2	0.8681	0.8764	0.6891		
	CISI [6]	0.8535	0.8599	0.6716		
	rSIM1 [11]	0.8574	0.8734	0.6865		
	rSIM2 [11]	0.7614	0.8089	0.5928		
	rSIM3 [11]	0.8621	0.8651	0.6778		
	rSIM4 [11]	0.8608	0.8653	0.6776		
	ESIM1 [12]	0.7818	0.8319	0.6357		
	ESIM2 [12]	0.8569	0.8452	0.6533		
	ESIM3 [12]	0.7638	0.8477	0.6558		
	ESIM4 [12]	0.7511	0.8583	0.6674		
	$CM^+7$ (best proposed)	0.9625	0.9608	0.8249		

**Table 4.** Comparison of results obtained for three major datasets using some combined metrics originally designed for singly distorted images with the "best" elementary metrics and the proposed methods. Performance of all metrics is expressed as Pearson, Spearman and Kendall correlation coefficients between the subjective quality scores and objective metrics. Better results from two alternatives are marked with bold font.

A relatively high performance of metrics optimized for the LIVEMD dataset applied for the MDID13 database is quite predictable since some of the images in both datasets are the same. Nevertheless, a good performance may be observed using the combined metrics developed for MDID for the images from the LIVEMD database. The MDID dataset—due to the highest number of images, diversity of distortions and the number of subjects who participated in experiments—may be considered as the most "demanding", hence, the combined metrics optimized for the other datasets do not outperform the use of the "best" elementary metric (IFS in this case). As the results of the cross-database validation of the CM<sup>+</sup> metrics have led to similar conclusions, they are not presented in the paper.

Nevertheless, from a practical point of view, a final recommendation of a "universal" combined metric suitable for all databases would be desired. Therefore, some additional experiments have been made using the "aggregated" correlation as the goal function. The "aggregated" correlation has been calculated as the weighted sum of four correlations computed for each dataset where their number of images has been used as the weight (before normalization), similarly as for the elementary metrics shown in Table 1.

Database		LIVEME	)		MDID13	5		MDID			MDIVL	
Metric	PCC	SROCC	KROCC									
CM2 <sub>LIVEMD</sub>	_	_	_	0.8234	0.8402	0.6391	0.8835	0.8853	0.7012	0.8411	0.8455	0.6494
CM3 <sub>LIVEMD</sub>	_	_	_	0.8334	0.8530	0.6525	0.8845	0.8837	0.7025	0.8510	0.8504	0.6547
CM4 <sub>LIVEMD</sub>	_	_	_	0.8531	0.8651	0.6674	0.8333	0.8351	0.6313	0.4798	0.5484	0.3791
CM5 <sub>LIVEMD</sub>	—	—	_	0.8527	0.8606	0.6675	0.8509	0.8472	0.6508	0.5639	0.6217	0.4819
CM6 <sub>LIVEMD</sub>	_	—	_	0.8538	0.8631	0.6675	0.8534	0.8493	0.6508	0.6364	0.6831	0.4819
CM2 <sub>MDID13</sub>	0.7194	0.6918	0.4822	_	_	_	0.7807	0.7581	0.5697	0.8281	0.8863	0.6986
CM3 <sub>MDID13</sub>	0.7423	0.7178	0.5075	_	_	-	0.7634	0.7387	0.5506	0.7978	0.8724	0.6734
CM4 <sub>MDID13</sub>	0.7325	0.7008	0.4910	_	—	_	0.8102	0.7981	0.6024	0.7706	0.8868	0.6986
CM5 <sub>MDID13</sub>	0.7419	0.7186	0.5170	_	_	_	0.8599	0.8579	0.6874	0.6581	0.7409	0.5475
CM6 <sub>MDID13</sub>	0.7517	0.7273	0.5170	_	_	_	0.8834	0.8809	0.6874	0.6825	0.7675	0.5475
CM2 <sub>MDID</sub>	0.7802	0.7619	0.5509	0.8415	0.8540	0.6543	_	_	_	0.8080	0.8539	0.6637
CM3 <sub>MDID</sub>	0.7876	0.7729	0.5636	0.8309	0.8435	0.6416	_	—	_	0.8041	0.8560	0.6667
CM4 <sub>MDID</sub>	0.7841	0.7639	0.5533	0.8279	0.8386	0.6355	_	—	_	0.7969	0.8497	0.6595
CM5 <sub>MDID</sub>	0.7835	0.7634	0.5533	0.8096	0.8219	0.6100	_	—	_	0.7917	0.8432	0.6253
CM6 <sub>MDID</sub>	0.7822	0.7626	0.5533	0.8046	0.8165	0.6100	_	—	_	0.7471	0.8243	0.6253
CM7 <sub>MDID</sub>	0.7817	0.7610	0.5511	0.7983	0.8108	0.6026	_	_	_	0.7470	0.8234	0.6250
CM2 <sub>MDIVL</sub>	0.6323	0.6642	0.4610	0.6922	0.7826	0.5855	0.7507	0.8599	0.6739	_	_	_
CM3 <sub>MDIVL</sub>	0.5449	0.6029	0.4128	0.5825	0.7390	0.5446	0.6966	0.8193	0.6269	_	_	_
CM4 <sub>MDIVL</sub>	0.5412	0.5949	0.4077	0.5855	0.7361	0.5418	0.6983	0.8146	0.6218	_	_	_
CM5 <sub>MDIVL</sub>	0.5166	0.5747	0.3961	0.5868	0.7384	0.5413	0.6698	0.8064	0.6141	_	_	_
CM6 <sub>MDIVL</sub>	0.5266	0.5815	0.3961	0.5926	0.7368	0.5413	0.6817	0.8084	0.6141	_	_	_
CM7 <sub>MDIVL</sub>	0.5275	0.5793	0.3944	0.5919	0.7358	0.5402	0.6827	0.8083	0.6141	_	_	_

**Table 5.** Results of the cross-database validation of the CM family of the combined metrics expressed by means of Pearson, Spearman, and Kendall correlation coefficients between the subjective quality scores and objective combined metrics. Better performance results than obtained for the best elementary metrics for each dataset are marked with bold font.

The results obtained for both proposed families of the combined metrics are presented in Table 6. It is worth noting that even considering all four databases, the correlations are higher than those achieved by the other combined metrics for single datasets as shown in Table 4. Analyzing the presented results, the advantages of the novel approach based on the weighted sum of metrics, leading to the  $CM^+$  family, may be observed for most metrics (better results from two alternatives are marked with bold font). Another interesting observation is that the "best" combinations of metrics in the  $CM^+$  family utilize different elementary metrics than in the case of the CM family. In some cases, due to the use of more parameters, it is also possible to achieve similar correlations using the  $CM^+$  approach with a smaller number of combined elementary metrics than using the CM family.

The graphical illustration of the correlation between the "best universal" combined metric CM<sup>+</sup>7 and subjective scores for individual datasets is provided in Figure 3, where the lowest correlation for LIVEMD may be easily observed. Nevertheless, due to the lowest number of images, this dataset may be considered as the least significant. Highly linear relationships between the subjective evaluation and objective metric achieved for three major datasets (PCC = 0.9387 for MDID, PCC = 0.8911 for MDID13, and PCC = 0.9122 for MDIVL, respectively, as shown over the plots in Figure 3) confirm the validity of the proposed approach. These results are still better in comparison to the results obtained for some alternative combined metrics presented in Table 4. The weights obtained for the elementary metrics that have different properties and various dynamic ranges, used in the CM<sup>+</sup>7 according to Formula (6), are provided in Table 7.



**Figure 3.** Scatter plots for the "best universal" elementary metric CM<sup>+</sup> obtained for each considered dataset together with the PCC values obtained for each dataset independently. Subjective quality scores are expressed as MOS and DMOS whereas CM<sup>+</sup> denotes the proposed objective combined metric.

**Table 6.** Performance of the "best" elementary, and "universal" CM and CM<sup>+</sup> metrics for all four databases in view of the aggregated (weighted) correlation with subjective scores. Better correlations from two families of the combined metrics are marked with bold font.

Metrics	PCC	Correlation Coefficients SROCC	KROCC	Denotation
VIF	0.8617	0.8817	0.7048	(elementary)
IFC, MCSD	0.8961	0.8975	0.7269	CM2
IFC, MCSD, FSIM	0.8998	0.9015	0.7322	CM3
IFC, MCSD, FSIM, MSVD	0.9019	0.9045	0.7362	CM4
IFC, MCSD, FSIM, MSVD, IW-PSNR	0.9027	0.9056	0.7369	CM5
IFC, MCSD, FSIM, MSVD, IW-PSNR, WSNR	0.9069	0.9118	0.7452	CM6
IFC, MCSD, FSIM, MSVD, IW-PSNR, WSNR, IFS	0.9095	0.9126	0.7467	CM7
PSIM, IFC	0.8956	0.9008	0.7297	CM <sup>+</sup> 2
PSIM, IFC, GMSD	0.9006	0.9039	0.7339	CM <sup>+</sup> 3
PSIM, IFC, GMSD, SIQAD	0.9051	0.9084	0.7395	$CM^+4$
PSIM, IFC, GMSD, SIQAD, SVQI	0.9091	0.9140	0.7459	$CM^+5$
PSIM, IFC, GMSD, SIQAD, SVQI, VIF	0.9121	0.9162	0.7498	$CM^+6$
PSIM, IFC, GMSD, SIQAD, SVQI, VIF, FSIM	0.9137	0.9178	0.7518	CM+7

Flomontary Motric O	Weights				
Elementary Metric Q <sub>i</sub>	a <sub>i</sub>	w <sub>i</sub>			
PSIM	3.7544	0.2552			
IFC	0.2027	$7.9157  imes 10^{-4}$			
GMSD	-0.8024	28.8528			
SIQAD	0.0678	2.5432			
SVQI	$1.8587  imes 10^{-5}$	-0.0013			
VIF	$-1.5179  imes 10^{-7}$	$7.0841 imes10^{-4}$			
FSIM	-0.0018	-0.0025			

**Table 7.** Weights obtained for the elementary metrics used in the proposed "best universal"  $CM^+7$  metric.

The conducted experiments have confirmed the hypothesis that the specificity of multiply distorted images requires a combination of different metrics since some of the previously proposed hybrid approaches have led to worse performance even in comparison to the "best" elementary metrics. Additionally, the application of the combination model proposed in the paper increases their performance meaningfully for most of the datasets considered in the paper as well as for all datasets treated as a whole. The application of the proposed approach makes it possible to improve both the quality prediction accuracy measured by the PCC and the prediction monotonicity reflected by both rank-order correlations (SROCC and KROCC).

### 6. Conclusions

Image quality assessment of multiply distorted images is still a challenging area of research as many elementary metrics designed using the IQA databases with singly distorted images have poor performance for multiple distorted ones. The application of the combined metrics makes it possible to increase the obtained performance; however, the results achieved using one of the available databases are not always directly applicable for the others. Therefore, our future research will concentrate on some other fusion strategies, including the use of genetic algorithms and neural networks for this purpose. Different approaches for feature extraction and network training are possible, however, as stated in the paper [34], *"the training set has to contain enough data samples to avoid overfitting"*. Meanwhile, even an application of relatively simple fusion models, as proposed in this paper, makes it possible to achieve much better results than may be achieved for a single metric.

Analyzing the results presented for the four available databases considered together, a significant increase of the aggregated correlation with subjective scores may be observed, not only in comparison to elementary metrics but also with the use of some other combined metrics, proposed earlier for images with single distortions. Those results confirm the practical usefulness and universality of the proposed approach, particularly the novel CM<sup>+</sup> metrics.

Since the proposed fusion model is not computationally demanding, its efficiency does not decrease significantly, assuming the possibility of parallel calculations of the elementary metrics. The only exception may be related to the memory limitations that would hinder the parallel computations of elementary metrics for large images. The time and memory requirements are dependent on the used hardware and the image size. For the parallel computation of metrics (e.g., 7 metrics for 8 independent threads), the calculation time of the final combined metric is nearly the same as for the "slowest" elementary metric being used.

The next step of research might be related to the application of the CNN-based metrics trained using the images affected by multiple distortions. Regardless of the different "nature" of the multiply distorted images compared to those affected by a single distortion, this direction of future research might be promising and will be considered. Nevertheless, its

significant limitation is the necessity of the development of some larger datasets containing multiply distorted images that may be used for training purposes.

Nevertheless, considering the presence of the multiple distortions in many electronic devices equipped with vision sensors, the proposed approach may be useful in various electronic systems used for image and video analysis purposes.

**Author Contributions:** Conceptualization, K.O. and V.V.L.; methodology, K.O. and V.V.L.; software, K.O.; validation, K.O. and P.L.; formal analysis, K.O. and V.V.L.; investigation, K.O.; resources, K.O. and V.V.L.; data curation, K.O. and P.L.; writing—original draft preparation, K.O.; writing—review and editing, K.O. and V.V.L.; visualization, K.O. and P.L.; project administration, K.O. and V.V.L.; funding acquisition, K.O. and V.V.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research is partially co-financed by the Polish National Agency for Academic Exchange (NAWA) and the Ministry of Education and Science of Ukraine under the project no. PPN/BUA/2019/1/00074 entitled "Methods of intelligent image and video processing based on visual quality metrics for emerging applications".

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

#### Abbreviations

The following abbreviations are used in this manuscript:

CISI	Combined Image Similarity Index
СМ	Combined Metric
CSSIM	Color Structural SIMilarity
CVSSI	Contrast and Visual Saliency Similarity-Induced Index
CW-SSIM	Complex Wavelet Structural SIMilarity
DCT	Discrete Cosine Transform
DSS	DCT Subbands Similarity
DMOS	Differential Mean Opinion Scores
ESIM	Evolutionary based Similarity Measure
FR	Full-Reference
FSIM	Feature SIMilarity
GMSD	Gradient Magnitude Similarity Deviation
GPU	Graphics Processing units
HaarPSI	Haar wavelet-based perceptual similarity metric
HVS	Human Visual System
ICA	Independent Component Analysis
IFC	Information Fidelity Criterion
IFS	independent feature similarity
IGM	internal generative mechanism
IQA	Image Quality Assessment
IW-PSNR	Information content weighted Peak Signal-to-Noise Ratio
IW-SSIM	Information content weighted Structural SIMilarity
JPEG	Joint Photographic Experts Group
KROCC	Kendall Rank Order Correlation Coefficient
LIVE	Laboratory for Image and Video Engineering
MCSD	Multiscale Contrast Similarity Deviation
MDID	Multiply Distorted Image Database
MDIVL	Multiply Distorted Imaging and Vision Laboratory database
MDSI	Mean Deviation Similarity Index
MOS	Mean Opinion Scores
MPEG	Moving Pictures Experts Group
MSE	Mean Square Error

MS-SSIM Multi-Scale Structural SIMilarity

18 c	of 20
------	-------

MC UNIOUE	Multi-model and Sharpness-weighted UNsupervised Image QUality
MIS-UNIQUE	Estimation
NR	No-Reference
NSS	Natural scene statistics
PCC	Pearson Linear Correlation Coefficient
PSIM	Perceptual SIMilarity
PSNR	Peak Signal-to-Noise Ratio
QILV	Quality Index based on Local Variance
RFSIM	Riesz-transform based Feature SIMilarity
RIQMC	reduced-reference image quality assessment of contrast change
RVSIM	Riesz transform and Visual contrast sensitivity-based feature SIMilarity
RR	Reduced-Reference
rSIM	Regression-based SIMilarity
SFF	Sparse Feature Fidelity
SROCC	Spearman Rank Order Correlation Coefficient
SR-SIM	Spectral Residual SIMilarity
SSIM	Structural SIMilarity
TID	Tampere Image Database
UNIQUE	UNsupervised Image QUality Estimation
UQI	Universal Image Quality Index
VIF	Visual Information Fidelity
VIFp	Pixel-domain Visual Information Fidelity
VSI	Visual Saliency-Induced Index
VSNR	Visual Signal-to-Noise Ratio
WASH	Wavelet Based Sharp Features

## References

- Athar, S.; Wang, Z. A comprehensive performance evaluation of image quality assessment algorithms. *IEEE Access* 2019, 7, 140030–140070. [CrossRef]
- 2. Chandler, D. Seven challenges in image quality assessment: Past, present, and future research. *ISRN Signal Process.* 2013, 2013, 905685. [CrossRef]
- 3. Niu, Y.; Zhong, Y.; Guo, W.; Shi, Y.; Chen, P. 2D and 3D image quality assessment: A survey of metrics and challenges. *IEEE Access* 2019, 7, 782–801. [CrossRef]
- 4. Zhai, G.; Min, X. Perceptual image quality assessment: A survey. Sci. China Inf. Sci. 2020, 63, 211301. [CrossRef]
- Okarma, K. Combined full-reference image quality metric linearly correlated with subjective assessment. In Artificial Intelligence and Soft Computing; Rutkowski, L., Scherer, R., Tadeusiewicz, R., Zadeh, L., Zurada, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; Volume 6113, pp. 539–546.
- 6. Okarma, K. Combined image similarity index. Opt Rev 2012, 19, 349–354. [CrossRef]
- Liu, T.J.; Lin, W.; Kuo, C.C.J. Image quality assessment using multi-method fusion. *IEEE Trans. Image Process.* 2013, 22, 1793–1807. [CrossRef]
- Lukin, V.; Ponomarenko, N.; Ieremeiev, O.; Egiazarian, K.; Astola, J. Combining full-reference image visual quality metrics by neural network. In *Human Vision and Electronic Imaging XX*; Rogowitz, B.E., Pappas, T.N., de Ridder, H., Eds.; SPIE: Bellingham, WA, USA, 2015; p. 93940K. [CrossRef]
- 9. Okarma, K.; Fastowicz, J.; Lech, P.; Lukin, V. Quality Assessment of 3D Printed Surfaces Using Combined Metrics Based on Mutual Structural Similarity Approach Correlated with Subjective Aesthetic Evaluation. *Appl. Sci.* **2020**, *10*, 6248. [CrossRef]
- 10. Ieremeiev, O.; Lukin, V.; Okarma, K.; Egiazarian, K. Full-Reference Quality Metric Based on Neural Network to Assess the Visual Quality of Remote Sensing Images. *Remote Sens.* **2020**, *12*, 2349. [CrossRef]
- 11. Oszust, M. A Regression-Based Family of Measures for Full-Reference Image Quality Assessment. *Meas. Sci. Rev.* 2016, 16, 316–325. [CrossRef]
- 12. Oszust, M. Decision Fusion for Image Quality Assessment using an Optimization Approach. *IEEE Signal Process. Lett.* **2016**, 23, 65–69. [CrossRef]
- 13. Ponomarenko, N.; Jin, L.; Ieremeiev, O.; Lukin, V.; Egiazarian, K.; Astola, J.; Vozel, B.; Chehdi, K.; Carli, M.; Battisti, F.; et al. Image database TID2013: Peculiarities, results and perspectives. *Signal Process. Image Commun.* **2015**, *30*, 57–77. [CrossRef]
- 14. Sun, W.; Zhou, F.; Liao, Q. MDID: A multiply distorted image database for image quality assessment. *Pattern Recognit.* 2017, *61*, 153–168. [CrossRef]
- 15. Wang, Z.; Bovik, A.C. A universal image quality index. IEEE Signal Process. Lett. 2002, 9, 81–84. [CrossRef]
- 16. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* 2004, 13, 600–612. [CrossRef] [PubMed]

- Wang, Z.; Simoncelli, E.P.; Bovik, A.C. Multiscale structural similarity for image quality assessment. In Proceedings of the 37th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 9–12 November 2003; pp. 1398–1402. [CrossRef]
- 18. Wang, Z.; Li, Q. Information content weighting for perceptual image quality assessment. *IEEE Trans. Image Process.* 2011, 20, 1185–1198. [CrossRef]
- Sampat, M.P.; Wang, Z.; Gupta, S.; Bovik, A.C.; Markey, M.K. Complex Wavelet Structural Similarity: A New Image Similarity Index. *IEEE Trans. Image Process.* 2009, 18, 2385–2401. [CrossRef] [PubMed]
- Zhang, L.; Zhang, L.; Mou, X.; Zhang, D. FSIM: A feature similarity index for image quality assessment. *IEEE Trans. Image Process.* 2011, 20, 2378–2386. [CrossRef] [PubMed]
- Aja-Fernandez, S.; Estepar, R.S.J.; Alberola-Lopez, C.; Westin, C.F. Image Quality Assessment based on Local Variance. In Proceedings of the 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, New York, NY, USA, 31 August–3 September 2006; pp. 4815–4818. [CrossRef]
- Ponomarenko, M.; Egiazarian, K.; Lukin, V.; Abramova, V. Structural Similarity index with predictability of image blocks. In Proceedings of the 17th International Conference on Mathematical Methods in Electromagnetic Theory (MMET), Kiev, Ukraine, 2–5 July 2018; pp. 115–118. [CrossRef]
- 23. Sheikh, H.R.; Bovik, A.C. Image information and visual quality. IEEE Trans. Image Process. 2006, 15, 430–444. [CrossRef] [PubMed]
- 24. Sheikh, H.R.; Bovik, A.C.; de Veciana, G. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Trans. Image Process.* 2005, 14, 2117–2128. [CrossRef]
- 25. Balanov, A.; Schwartz, A.; Moshe, Y.; Peleg, N. Image quality assessment based on DCT subband similarity. In Proceedings of the International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27 September 2015; pp. 2105–2109. [CrossRef]
- Dumic, E.; Grgic, S.; Grgic, M. IQM2: New image quality measure based on steerable pyramid wavelet transform and structural similarity index. SIViP 2014, 8, 1159–1168. [CrossRef]
- 27. Wu, J.; Lin, W.; Shi, G.; Liu, A. Perceptual quality metric with internal generative mechanism. *IEEE Trans. Image Process.* 2013, 22, 43–54. [CrossRef] [PubMed]
- Chang, H.W.; Zhang, Q.W.; Wu, Q.G.; Gan, Y. Perceptual image quality assessment by independent feature detector. *Neurocomputing* 2015, 151, 1142–1152. [CrossRef]
- Gu, K.; Li, L.; Lu, H.; Min, X.; Lin, W. A fast reliable image quality predictor by fusing micro- and macro-structures. *IEEE Trans. Ind. Electron.* 2017, 64, 3903–3912. [CrossRef]
- Chang, H.W.; Yang, H.; Gan, Y.; Wang, M.H. Sparse Feature Fidelity for perceptual image quality assessment. *IEEE Trans. Image Process.* 2013, 22, 4007–4018. [CrossRef]
- Temel, D.; Prabhushankar, M.; AlRegib, G. UNIQUE: Unsupervised Image Quality Estimation. *IEEE Signal Process. Lett.* 2016, 23, 1414–1418. [CrossRef]
- Prabhushankar, M.; Temel, D.; AlRegib, G. MS-UNIQUE: Multi-model and Sharpness-weighted Unsupervised Image Quality Estimation. *Electron. Imaging* 2017, 2017, 30–35. [CrossRef]
- 33. Bosse, S.; Maniry, D.; Muller, K.R.; Wiegand, T.; Samek, W. Neural network-based full-reference image quality assessment. In Proceedings of the 2016 Picture Coding Symposium (PCS), Nuremberg, Germany, 4–7 December 2016; pp. 1–5. [CrossRef]
- Bosse, S.; Maniry, D.; Muller, K.R.; Wiegand, T.; Samek, W. Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment. *IEEE Trans. Image Process.* 2018, 27, 206–219. [CrossRef] [PubMed]
- 35. Nafchi, H.Z.; Shahkolaei, A.; Hedjam, R.; Cheriet, M. Mean Deviation Similarity Index: Efficient and reliable full-reference image quality evaluator. *IEEE Access* 2016, *4*, 5579–5590. [CrossRef]
- Xue, W.; Zhang, L.; Mou, X.; Bovik, A.C. Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index. *IEEE Trans. Image Process.* 2014, 23, 684–695. [CrossRef] [PubMed]
- 37. Reisenhofer, R.; Bosse, S.; Kutyniok, G.; Wiegand, T. A Haar wavelet-based perceptual similarity index for image quality assessment. *Signal Process. Image Commun.* **2018**, *61*, 33–43. [CrossRef]
- 38. Yang, G.; Li, D.; Lu, F.; Liao, Y.; Yang, W. RVSIM: A feature similarity method for full-reference image quality assessment. *J. Image Video Proc.* 2018, 2018, 6. [CrossRef]
- Zhang, L.; Zhang, L.; Mou, X. RFSIM: A feature based image quality assessment metric using Riesz transforms. In Proceedings of the 2010 IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010; pp. 321–324. [CrossRef]
- 40. Jia, H.; Zhang, L.; Wang, T. Contrast and Visual Saliency Similarity-Induced index for assessing image quality. *IEEE Access* 2018, *6*, 65885–65893. [CrossRef]
- Cheraaqee, P.; Mansouri, A.; Mahmoudi-Aznaveh, A. Incorporating gradient direction for assessing multiple distortions. In Proceedings of the 4th International Conference on Pattern Recognition and Image Analysis (IPRIA), Tehran, Iran, 6–7 March 2019; pp. 109–113. [CrossRef]
- 42. Miao, X.; Chu, H.; Liu, H.; Yang, Y.; Li, X. Quality assessment of images with multiple distortions based on phase congruency and gradient magnitude. *Signal Process. Image Commun.* **2019**, *79*, 54–62. [CrossRef]
- Mitsa, T.; Varkur, K. Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms. In Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing, Minneapolis, MN, USA, 27–30 April 1993; Volume 5, pp. 301–304. [CrossRef]

- Ponomarenko, N.; Ieremeiev, O.; Lukin, V.; Egiazarian, K.; Carli, M. Modified image visual quality metrics for contrast change and mean shift accounting. In Proceedings of the 2011 11th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM), Polyana, Ukraine, 23–25 February 2011; pp. 305–311.
- Chandler, D.; Hemami, S. VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images. *IEEE Trans. Image Process.* 2007, 16, 2284–2298. [CrossRef] [PubMed]
- Zhang, L.; Shen, Y.; Li, H. VSI: A Visual Saliency-Induced Index for Perceptual Image Quality Assessment. *IEEE Trans. Image Process.* 2014, 23, 4270–4281. [CrossRef] [PubMed]
- 47. Wang, T.; Zhang, L.; Jia, H.; Li, B.; Shu, H. Multiscale contrast similarity deviation: An effective and efficient index for perceptual image quality assessment. *Signal Process. Image Commun.* **2016**, *45*, 1–9. [CrossRef]
- Zhang, L.; Li, H. SR-SIM: A fast and high performance IQA index based on spectral residual. In Proceedings of the 2012 19th IEEE International Conference on Image Processing, Orlando, FL, USA, 30 September–3 October 2012; pp. 1473–1476. [CrossRef]
- Reenu, M.; David, D.; Raj, S.S.A.; Nair, M.S. Wavelet Based Sharp Features (WASH): An Image Quality Assessment Metric Based on HVS. In Proceedings of the 2013 2nd International Conference on Advanced Computing, Networking and Security, Mangalore, India, 15–17 December 2013; pp. 79–83. [CrossRef]
- 50. Xia, Z.; Gu, K.; Wang, S.; Liu, H.; Kwong, S. Toward Accurate Quality Estimation of Screen Content Pictures With Very Sparse Reference Information. *IEEE Trans. Ind. Electron.* 2020, 67, 2251–2261. [CrossRef]
- Ni, Z.; Ma, L.; Zeng, H.; Cai, C.; Ma, K.K. Gradient Direction for Screen Content Image Quality Assessment. *IEEE Signal Process*. *Lett.* 2016, 23, 1394–1398. [CrossRef]
- 52. Gu, K.; Zhai, G.; Lin, W.; Liu, M. The Analysis of Image Contrast: From Quality Assessment to Automatic Enhancement. *IEEE Trans. Cybern.* **2016**, *46*, 284–297. [CrossRef]
- Jayaraman, D.; Mittal, A.; Moorthy, A.K.; Bovik, A.C. Objective quality assessment of multiply distorted images. In Proceedings of the 46th Asilomar Conference on Signals, Systems and Computers (ASILOMAR), Pacific Grove, CA, USA, 4–7 November 2012. [CrossRef]
- 54. Gu, K.; Zhai, G.; Yang, X.; Zhang, W. Hybrid no-reference quality metric for singly and multiply distorted images. *IEEE Trans. Broadcast.* **2014**, *60*, 555–567. [CrossRef]
- Corchs, S.; Gasparini, F. A multidistortion database for image quality. In *Computational Color Imaging. CCIW 2017*; Bianco, S., Schettini, R., Trémeau, A., Tominaga, S., Eds.; Springer: Berlin/Heidelberg, Germany, 2017; Volume 10213, pp. 95–104.\_8. [CrossRef]
- Ghadiyaram, D.; Bovik, A.C. Massive Online Crowdsourced Study of Subjective and Objective Picture Quality. *IEEE Trans. Image Process.* 2016, 25, 372–387. [CrossRef] [PubMed]