CrossMark

Image quality assessment with lasso regression and pairwise score differences

Mariusz Oszust[1] (ORCID)

Abstract The reception of multimedia applications often depends on the quality of processed and displayed visual content. This is the main reason for the development of automatic image quality assessment (IQA) techniques which try to mimic properties of human visual system and produce objective scores for evaluated images. Most of them require a training step in which subjective scores, obtained in tests with human subjects, are used for parameters tuning. In this paper, it is shown that pairwise score differences (PSD) can be successfully used for training a full-reference hybrid IQA measure based on the least absolute shrinkage and selection operator (lasso) regression. The results of extensive experimental evaluation on four largest IQA benchmarks show that the proposed IQA technique is statistically better than its version trained using raw scores, and both approaches are statistically better than state-of-the-art full-reference IQA measures. They are also better than other hybrid approaches. In the paper, the evaluation protocol is extended with tests using PSD.

Keywords Image quality assessment · Full-reference · Lasso regression · Evaluation

1 Introduction

With the rapid development of different image processing methods, a considerable attention is given to techniques which are trying to mimic human visual perception [9, 46]. Image quality assessment (IQA) techniques measure the quality of presented images, and often support, or compare, image enhancement, restoration, or denoising techniques [3, 5, 29, 32].

✉ Mariusz Oszust
marosz@kia.prz.edu.pl
http://marosz.kia.prz.edu.pl

[1] Department of Computer and Control Engineering, Rzeszow University of Technology, W. Pola 2, 35-959 Rzeszow, Poland

According to the availability of a reference image, IQA measures are divided into full-reference, reduced-reference, and no-reference techniques [4, 19, 31, 47].

This paper focuses on the full-reference IQA measures. Over the last decade, many different full-reference IQA approaches have been introduced [4], starting from simple peak signal-to-noise ratio (PSNR), or noise quality measure (NQM) [8] in which a linear frequency distortion and an additive noise injection are used for the distorted image modelling. The popular structural similarity (SSIM) measure [41], being the extension of universal image quality index (UQI) [39], uses loss of correlation, luminance distortion, and contrast distortion. SSIM was further extended using a multi-scale approach (MSSIM) [41] or statistical models of natural scenes, as it can be seen in information content weighted SSIM (IWSSIM) [40]. Such statistical models are also utilised in information fidelity criterion (IFC) [33] and visual information fidelity (VIF) [34]. Feature similarity index (FSIM) [51] and its version for colour images (FSIMc) employ phase congruency and image gradient magnitude. In Riesz-transform based feature similarity measure (RFSIM) [50] Riesz-transform features at key locations between the distorted image and its reference image are used. SURF-SIM [38], in turn, uses Speed Up Robust Features (SURF) in order to detect multiscale differences between features. Spectral residual based similarity measure (SRSIM) [49] and visual saliency-induced index (VSI) [52] use visual saliency maps. Contrast changes and luminance distortions are used in gradient similarity (GSM) measure [16], and inter-patch and intra-patch similarities were modelled in [54] using modified normalised correlation coefficient and image curvature. The edge based image quality assessments (EBIQA) measure is based on different edge features extracted from distorted image and its pristine equivalent [1]. In [11], a novel pooling strategy based on the harmonic mean was proposed.

In the literature, there are also approaches in which several IQA techniques are combined into a hybrid, or a fusion, measure. For example, in the most apparent distortion algorithm (MAD) [14] local luminance and contrast masking evaluate high-quality images, while changes in the local statistics of spatial-frequency components are used for images with a low quality. Information obtained using saliency maps, gradient and contrast information was fused in [30]. In [21, 22], scores of MSSIM, VIF and R-SVD were non-linearly combined. A preliminary work with non-linear combination of several IQA measures selected by a genetic algorithm was shown in [23]. In [17], SNR, SSIM, VIF, and VSNR were combined using canonical correlation analysis, and a regularised regression was used to combine up to seven IQA models in [13]. In [25], a support vector machines classifier used for predicting of the distortion followed by a fusion of SSIM, VSNR, and VIF using k-nearest-neighbour regression, was proposed. An adaptive combination of IQA approaches with an edge-quality based on preservation of edge direction was introduced in [26]. In [18], a fusion measure using a support vector regression approach was proposed. Lukin et al. [20] introduced a fusion measure which combines six IQA measures using a neural network. In [48], kernel ridge regression was used to combine found perceptually meaningful structures and local distortion measurements. In other approaches, adaptive weighting [2] or internal generative mechanism [43] were considered in order to obtain hybrid measures.

For evaluation of IQA approaches, specific IQA benchmark databases have been introduced [14, 27, 28, 42]. They contain pristine images, their distorted equivalents and subjective human evaluations in the form of mean opinion scores (MOS) or differential MOS (DMOS). Some images from these benchmarks with subjective scores are often used for tuning parameters of many developed methods, e.g., [13, 20, 44, 52, 54]. Here, the number of used images should be small in order to obtain the benchmark-independent solution. In this paper, a novel full-reference hybrid IQA measure is proposed which employs

regularized least-squares regression using the least absolute shrinkage and selection operator (lasso) [36, 37]. This technique combines objective scores produced by up to 16 full-reference IQA measures. The lasso regression was applied since it performs selection of the most important predictors, what makes the usage of such combined measure more practical. Finally, only several IQA measures take part in the fusion. In the proposed approach, the regression coefficients are determined using part of images and their scores from benchmark databases. It is shown that the proposed hybrid measure is significantly better if pairwise score differences (PSD) are used instead of raw scores. These differences can be also used for performance evaluation of IQA measures. The application of PSD can be motivated by the organisation of some IQA tests with human subjects [28], where the observer compares distorted images with each other, taking into account the pristine image. It can be assumed that PSD can also be used in the development of IQA measures that require supervised learning. The hybrid measures developed using raw scores or PSD in the lasso regression are compared with the state-of-the-art techniques on four largest IQA benchmark image datasets using well-established evaluation protocol, as well as statistical significance tests.

The rest of this paper is organised as follows. Section 2 presents the proposed hybrid IQA measure. In Section 3, the approach is compared with state-of-the-art measures using four IQA benchmarks, and, finally concluding remarks are presented in Section 4.

2 Proposed approach

Let $Q_1, \ldots, Q_M$ be the objective scores of M IQA measures seen as predictor variables in multiple linear regression model [36]. In the model, S^o is the estimated response, or objective score, of the resulted hybrid IQA measure. It can be written as follows:

$$S^o = B_0 + \sum_{m=1}^{M} Q_m B_m + \epsilon, \tag{1}$$

where B contains fitted coefficients estimated by minimising the mean squared difference between the outcome, i.e., the vector of subjective scores S^s, and predicted outcome, S^o; ϵ represents a relationship between Q and S^s which is rejected from the equation.

For large number of predictors, it would be desirable to select those which are the most informative. This also leads to a more practical IQA hybrid approach, consisting of only several IQA measures. One of possible approaches to the problem of predictor selection is to use a penalised regression in the lasso form [36]. In the regression, for a given λ, the lasso determines B solving the following optimisation problem:

$$\min_{B} \left(\frac{1}{2} \sum_{n=1}^{N} (S_n^o - B_0 - \sum_{m=1}^{M} Q_{nm} B_m)^2 + \lambda \sum_{m=1}^{M} |B_m| \right), \tag{2}$$

where N is the number of objective scores, and λ is a regularization parameter. In other words, the lasso minimises the residual sum of squares with the constant α:

$$\sum_{m=1}^{M} |B_m| \leq \alpha. \tag{3}$$

In the approach, λ value which minimises the mean squared error was used to determine coefficients.

In experiments, the following $M = 16$ publicly available full-reference IQA measures were used: VSI [52], FSIM [51], FSIMc [51], GSM [16], IFC [33], IW-SSIM [40], MAD [14], MSSIM [41], NQM [8], PSNR [35], RFSIM [50], SR-SIM [49], SSIM [42], VIF [34], IFS [7], and SFF [6]. They were used for assessment of processed images and then PSD were obtained. Most of these approaches present state-of-the-art performance, and their inclusion was mainly influenced by the need of achieving a broad sample of various approaches mimicking human visual system. It is assumed that the lasso regression would be able to select several IQA measures and develop the well-performing hybrid measure.

The proposed approach uses first 20 % images and their subjective scores from a given benchmark dataset, in order to obtain the regression coefficients. There are four largest IQA benchmark image datasets; therefore, four hybrid measures are introduced. In the literature, different numbers of images with scores were used for this purpose, ranging from 20 % [38], through 30 % [52], 50 % [48], and 100 % [13, 26, 44], to several datasets jointly [54].

In experiments, the following four largest IQA benchmarks were used: TID2013 [28], TID2008 [27], CSIQ [14], and LIVE [42]. The number of images in each benchmark, as well as the number of distortions and their levels are shown in Table 1. Since the number of learning images in the subset is small, the number of scores used in the regression can be considerably increased by employing PSD. To the best knowledge of the author, PSD have not been used for training of IQA measures. In this paper, the lasso regression produces the hybrid IQA measure trained with the small subset of images and scores obtained for $M = 16$ IQA measures, as well as trained with pairwise differences of these scores. The obtained fitted coefficients, $\boldsymbol{B}$, indicate the number and contribution of the most informative IQA measures. Only these measures were used in the quality assessment of the test images. For a selected reference image, all score differences between its distorted equivalents are calculated. For example, for 5 images with 24 distortions and 5 distortion levels, 600 images and scores are available in the typical learning scenario, or, as it is introduced in this paper, $\sum_{k=1}^{5} \binom{120}{2} = 7140$ pairwise score differences for these images can be used. It is assumed that only scores of distorted images that have the same reference image are compared. The usage of PSD can be also motivated by the tristimulus methodology for performing tests with human observers [28], in which two distorted images are presented with their pristine equivalent at the same time. Then, the observer selects which distorted image has the better quality, what requires evaluation of each distorted image separately, looking at the pristine image, and jointly, while making the decision on their relative quality. Such pairwise image comparison is used to determine subjective opinions for assessed images [28].

In the experiments, images from a given benchmark dataset were divided into five disjoint subsets. There are 20 % of all images in each subset, and each image was evaluated by 16 IQA measures. Finally, after the application of the proposed approach, 40 hybrid IQA

Table 1 IQA benchmark image datasets

Benchmark	No. of reference images	No. of distorted images	No. of distortions
TID2013 [28]	25	3000	24
TID2008 [27]	25	1700	17
CSIQ [14]	30	866	6
LIVE [42]	29	779	5

measures were obtained; half of them was trained on PSD. For convenience of presentation, only measures, namely *lasso regression SImilarity Measures* (lrSIMs), obtained for the first 20 % of benchmark images are written as follows:

$$lrSIM_1^{1a} = 10.214VSI - 1.5221MAD - 0.5705PSNR$$
$$+ 0.7827RFSIM + 0.5723VIF + 1.9253IFS \tag{4}$$

$$lrSIM_1^{2a} = 8.2432VSI - 2.9136MAD - 1.0000PSNR$$
$$+ 1.0432VIF + 1.8354IFS \tag{5}$$

$$lrSIM_2^{1a} = 0.5107VSI - 1.5079MAD + 0.5439PSNR$$
$$+ 1.1451RFSIM + 0.3124SRSIM + 1.0850VIF$$
$$+ 0.6202IFS + 5.7429SFF \tag{6}$$

$$lrSIM_2^{2a} = -2.5348MAD + 0.6056RFSIM + 1.6761SRSIM$$
$$+ 1.3234VIF + 0.8086IFS + 3.8507SFF \tag{7}$$

$$lrSIM_3^{1a} = 0.3887MAD - 0.1408RFSIM - 0.1969VIF \tag{8}$$

$$lrSIM_3^{2a} = 0.5193MAD - 0.2754VIF - 0.0543IFS \tag{9}$$

$$lrSIM_4^{1a} = 14.913IFC + 72.26MAD + 1.5549NQM$$
$$+ 2.5175PSNR + 20.989SRSIM$$
$$- 36.315SSIM - 43.421VIF \tag{10}$$

$$lrSIM_4^{2a} = 11.906GSM + 6.8190IWSSIM + 71.034MAD$$
$$+ 6.0730MSSIM - 38.154VIF - 15.709IFS \tag{11}$$

In the (4)–(11), the number in the subscript denotes the benchmark whose part of images were used for the development of the measure, 1 for TID2013, 2, 3, 4 for TID2008, CSIQ, and LIVE, respectively; the number in superscript, in turn, indicates that the measure was developed using scores "1" or PSD "2"; the letter in the subscript denotes which subset of training images was used (five letters: *a-e*). In the evaluation (see Section 3), results for the *a* subset or for all subsets together in the form of the mean value are reported. Taking into account all obtained hybrid measures, it can be said that one hybrid measure uses 5.725 single IQA measures, on average. Among the mostly used IQA measures, VIF was selected 40 times, MAD 39, IFS 29, FSIM 25, RFSIM 22, PSNR 20, SFF 19, and VSI 17. The remaining IQA measures were used less than 15 times each. Interestingly, FSIMc was not used at all, and NQM, SSIM, or MSSIM were used less than five times, what can indicate that their features were replaced by the remaining IQA measures. Some measures contributed more than others, what is reflected by the weights. For example, in hybrid measures (4)–(5) VSI was the most contributing technique, in (6)–(7) MAD with SFF contributed more than other techniques, and in (8)–(11) MAD with VIF. The sign the weight mostly depends on the sign of correlation between the objective scores produced by the measure and the subjective scores in the benchmark. Experiments were performed using Matlab 7.14 with Statistics Toolbox.

3 Experimental evaluation

According to the widely-used protocol [10, 35], IQA measures are compared with each other using the following performance indices: Spearman Rank order Correlation Coefficient (SRCC), Kendall Rank order Correlation Coefficient (KRCC), Pearson linear

Correlation Coefficient (PCC) and Root Mean Square Error (RMSE). The prediction monotonicity is evaluated by SRCC and KRCC, and the prediction accuracy by PCC and RMSE. These performance indices are calculated after a non-linear mapping between a vector of objective scores, S^o, and subjective scores, MOS or DMOS, denoted here by S^s, using the following mapping function for the non-linear regression [35]:

$$S_m^o = \beta_1 \left(\frac{1}{2} - \frac{1}{exp(\beta_2(S^o - \beta_3))} \right) + \beta_4 S^o + \beta_5, \qquad (12)$$

where $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_5]$ are parameters of the non-linear regression model [35], and S_m^o is the non-linearly mapped S^o. PCC and RMSE use S^o:

$$PCC(S_m^o, S^s) = \frac{\bar{S}_m^{o\,T} \bar{S}^s}{\sqrt{\bar{S}_m^{o\,T} \bar{S}_m^o \bar{S}^{s\,T} \bar{S}^s}}, \qquad (13)$$

where $\bar{S}_m^o$ and $\bar{S}^s$ denote the mean-removed vectors.

$$RMSE(S_m^o, S^s) = \sqrt{\frac{(S_m^o - S^s)^T(S_m^o - S^s)}{m}}. \qquad (14)$$

SRCC is calculated as:

$$SRCC(S^o, S^s) = 1 - \frac{6 \sum_{i=1}^m d_i^2}{m(m^2 - 1)}, \qquad (15)$$

where d_i is the difference between i^{th} image in $\hat{S}$ and S, m is the total number of images. In KRCC, the number of concordant pairs in the dataset, m_c, is used, as well as the number of discordant pairs, m_d:

$$KRCC(S^o, S^s) = \frac{m_c - m_d}{0.5m(m - 1)}. \qquad (16)$$

The values of RMSE closer to 0 are considered better, in contrary to SRCC, KRCC, and PCC whose values should be close to 1.

Table 2 contains mean values of evaluation indices for all developed hybrid measures on four benchmarks. There are 178500 values of PSD for TID2013, 56950, 12068, and 10081, for TID2008, CSIQ, and LIVE, respectively. It can be seen that within the benchmark, mean values for hybrid measures trained with PSD, indicated with "2" in the superscript (e.g., lrSIM$_{1-4}^2$), were in almost all cases better than hybrid measures learned using images from the benchmark and their raw scores, lrSIM$_{1-4}^1$. This indicates that PSD carry more information than raw scores.

Extension of the typical testing protocol with results obtained using PSD may lead to more quantitative conclusions about the performance of evaluated IQA measures. Therefore, the proposed hybrid measures, trained using PSD and represented by lrSIM$_{1-4}^{2a}$, are compared with state-of-the-art IQA measures using four performance indices calculated with raw scores, and with PSD. The results of comparison are shown in Tables 3 and 4. The overall results for RMSE do not take into account LIVE dataset due to the range difference, and weighted results were obtained using the number of images in the benchmark as its weight. The tables contain the six best IQA measures that were considered in the regression, out of 16, and lrSIM$_{1-4}^{2a}$, the four best performing measures for each evaluation index are written in bold. The obtained results reveal that all presented lrSIMs clearly outperformed compared IQA measures. For TID2013, where VSI was the best performing non-hybrid measure, hybrid measures trained on images from benchmark with considerably

Table 2 Comparison of mean values of RMSE for developed hybrid measures on four IQA benchmark datasets using raw scores or PSD

$lrSIM_1^1$	$lrSIM_1^2$	$lrSIM_2^1$	$lrSIM_2^2$	$lrSIM_3^1$	$lrSIM_3^2$	$lrSIM_4^1$	$lrSIM_4^2$
TID2013							
0.5266	0.5225	0.6092	0.6056	0.6203	0.6333	0.6709	0.6412
TID2013, PSD							
0.7150	0.6910	0.8487	0.8252	0.9412	0.8899	0.9845	0.9090
TID2008							
0.5794	0.5794	0.5532	0.5518	0.5994	0.5917	0.6677	0.6190
TID2008, PSD							
0.8015	0.7813	0.7467	0.7385	0.8158	0.7856	0.9116	0.8187
CSIQ							
0.0735	0.0690	0.0659	0.0638	0.0679	0.0605	0.0766	0.0639
CSIQ, PSD							
0.1133	0.1041	0.0912	0.0892	0.0887	0.0767	0.1003	0.0807
LIVE							
7.1739	6.9237	6.7223	6.5120	6.6943	6.2567	6.9266	6.1925
LIVE, PSD							
13.835	13.142	11.040	10.951	9.1425	8.9324	9.8535	8.1755

The measures developed with PSD ($lrSIM_{1\text{-}4}^2$) are better than measures developed using raw scores ($lrSIM_{1\text{-}4}^1$)

less number of distortions that are present in TID2013, i.e., on CSIQ and LIVE, performed worse than measures trained on TID benchmarks. IQA measures that can be seen in models trained on CSIQ and LIVE, which share most of distortion types, are performing worse on newly introduced TID benchmarks, and that can explain worse results for $lrSIM_3^{2a}$ and $lrSIM_4^{2a}$. However, these results are still better, taking into account overall performance, than for other non-hybrid IQA measures. Weighted results seem to favour IQA approaches performing better on TID2013 due to large number of images used as its weight. However, weighted results for tests with PSD show superior performance of all introduced hybrid measures. Interestingly, evaluation results on benchmarks with PSD, seen in Table 4, allow further assessment of the performance of compared non-hybrid IQA measures. There are cases in which some measures were better in this test than it was shown in the typical evaluation with raw scores. For example, for TID2008 with PSD, SFF was better than IFS and VSI, while in the previous evaluation, their precedence was reversed, with VSI as the leading technique. Interestingly, SFF was introduced before IFS by the same authors. Furthermore, for the results on CSIQ with PSD, MAD clearly outperformed newly introduced SFF and IFS, what was not evident in the known evaluation. MAD's performance was also confirmed in tests on LIVE with PSD, where it was the fourth best IQA measure.

The evaluation results on benchmark datasets showed superior performance of introduced family of hybrid IQA measures, lrSIMs. However, it would be desirable to prove that the approach is statistically better. In statistical significance tests, hypothesis tests based on the prediction residuals of each measure after non-linear mapping were conducted using F-test [14], where the smaller residual variance denoted the better prediction. The results of statistical significance tests on LIVE benchmark are presented in Fig. 1. The tests cover all 16 IQA measures that were considered in the regression. In the figure, the number "1",

Table 3 Comparison of hybrid measures with the six best state-of-the-art IQA measures on four benchmark datasets

	VSI	FSIM	MAD	SRSIM	IFS	SFF	lrSIM_1^{2a}	lrSIM_2^{2a}	lrSIM_3^{2a}	lrSIM_4^{2a}
TID2013										
SRCC	**0.8965**	0.8015	0.7807	0.7999	**0.8697**	**0.8513**	**0.8850**	0.8211	0.8070	0.8082
KRCC	**0.7183**	0.6289	0.6035	0.6314	**0.6785**	**0.6581**	**0.7060**	0.6463	0.6298	0.6307
PCC	**0.9000**	0.8589	0.8267	0.8590	**0.8791**	**0.8706**	**0.9012**	0.8676	0.8604	0.8604
RMSE	**0.5404**	0.6349	0.6975	0.6347	**0.5909**	**0.6099**	**0.5372**	0.6164	0.6318	0.6317
TID2008										
SRCC	**0.8979**	0.8805	0.8340	0.8913	0.8903	0.8767	**0.9013**	**0.9073**	0.8916	**0.8923**
KRCC	**0.7123**	0.6946	0.6445	**0.7149**	0.7009	0.6882	**0.7217**	**0.7332**	0.7083	0.7089
PCC	0.8762	0.8738	0.8306	0.8866	0.8810	0.8817	**0.9009**	**0.9100**	**0.8943**	**0.8917**
RMSE	0.6466	0.6525	0.7473	0.6206	0.6349	0.6333	**0.5824**	**0.5563**	**0.6005**	**0.6073**
CSIQ										
SRCC	0.9423	0.9242	0.9466	0.9319	0.9582	0.9627	**0.9630**	**0.9688**	**0.9670**	**0.9681**
KRCC	0.7857	0.7567	0.7970	0.7725	0.8165	0.8288	**0.8302**	**0.8449**	**0.8423**	**0.8456**
PCC	0.9279	0.9120	0.9500	0.9250	0.9576	0.9643	**0.9680**	**0.9722**	**0.9720**	**0.9727**
RMSE	0.0979	0.1077	0.0820	0.0997	0.0757	0.0695	**0.0659**	**0.0615**	**0.0616**	**0.0610**
LIVE										
SRCC	0.9524	0.9634	0.9669	0.9618	0.9599	0.9649	**0.9691**	**0.9730**	**0.9732**	**0.9735**
KRCC	0.8058	0.8337	0.8421	0.8299	0.8254	0.8365	**0.8432**	**0.8546**	**0.8554**	**0.8567**
PCC	0.9482	0.9597	0.9675	0.9553	0.9586	0.9632	**0.9690**	**0.9723**	**0.9735**	**0.9741**
RMSE	8.6816	7.6781	6.9073	8.0813	7.7765	7.3461	**6.7458**	**6.3846**	**6.2452**	**6.1742**
Overall direct										
SRCC	**0.9223**	0.8924	0.8821	0.8962	**0.9195**	0.9139	**0.9296**	**0.9176**	0.9097	0.9105
KRCC	0.7555	0.7285	0.7218	0.7372	0.7553	0.7529	**0.7753**	**0.7697**	**0.7590**	**0.7605**
PCC	0.9131	0.9011	0.8937	0.9065	0.9191	0.9200	**0.9348**	**0.9305**	**0.9251**	**0.9247**
RMSE	**0.4283**	0.4650	0.5089	0.4517	0.4338	0.4376	**0.3952**	**0.4114**	**0.4313**	0.4333
Overall weighted										
SRCC	**0.9102**	0.8598	0.8412	0.8628	**0.8988**	**0.8877**	**0.9107**	0.8836	0.8725	0.8735
KRCC	**0.7370**	0.6898	0.6711	0.6981	**0.7220**	0.7121	**0.7446**	**0.7231**	0.7084	0.7096
PCC	**0.9036**	0.8828	0.8625	0.8876	**0.9004**	0.8981	**0.9189**	**0.9065**	0.8991	0.8986
RMSE	**0.5025**	0.5566	0.6150	0.5456	**0.5226**	0.5313	**0.4762**	**0.5101**	0.5319	0.5337

In the table, raw subjective scores were used

The four best IQA measures for each evaluation index are written in bold. Overall results for RMSE do not take into account LIVE benchmark, since they are ranged differently than for the other benchmarks. For weighted results, the number of images in the benchmark was used as its weight

"0" or "-1" in the cell denotes that the measure in the row is statistically better with the confidence greater than 95 %, indistinguishable, or worse than the measure in the column,

Table 4 Comparison of hybrid measures with the six best state-of-the-art IQA measures on four benchmark datasets

	VSI	FSIM	MAD	SRSIM	IFS	SFF	lrSIM_1^{2a}	lrSIM_2^{2a}	lrSIM_3^{2a}	lrSIM_4^{2a}
	TID2013									
SRCC	**0.8885**	0.8489	0.8203	0.8447	**0.8685**	**0.8651**	**0.9116**	0.8604	0.8420	0.8423
KRCC	**0.7100**	0.6660	0.6459	0.6616	**0.6834**	0.6786	**0.7487**	**0.6893**	0.6694	0.6689
PCC	**0.8801**	0.8585	0.8314	0.8530	0.8658	**0.8673**	**0.9185**	**0.8770**	0.8557	0.8557
RMSE	**0.8350**	0.9017	0.9771	0.9177	0.8799	**0.8753**	**0.6952**	**0.8451**	0.9099	0.9101
	TID2008									
SRCC	0.8655	0.8695	0.8535	0.8786	0.8684	0.8756	**0.9155**	**0.9193**	**0.9060**	**0.9059**
KRCC	0.6780	0.6829	0.6718	0.6948	0.6812	0.6896	**0.7506**	**0.7575**	**0.7364**	**0.7358**
PCC	0.8492	0.8629	0.8550	0.8714	0.8632	0.8761	**0.9160**	**0.9221**	**0.9082**	**0.9077**
RMSE	1.0086	0.9650	0.9904	0.9369	0.9641	0.9206	**0.7659**	**0.7388**	**0.7991**	**0.8014**
	CSIQ									
SRCC	0.8853	0.8544	0.9543	0.8478	0.9348	0.9408	**0.9665**	**0.9735**	**0.9761**	**0.9773**
KRCC	0.7188	0.6837	0.8167	0.6750	0.7811	0.7897	**0.8438**	**0.8607**	**0.8669**	**0.8704**
PCC	0.8874	0.8652	0.9573	0.8664	0.9316	0.9346	**0.9672**	**0.9743**	**0.9766**	**0.9775**
RMSE	0.1674	0.1821	0.1050	0.1814	0.1320	0.1292	**0.0923**	**0.0818**	**0.0781**	**0.0766**
	LIVE									
SRCC	0.9031	0.9273	**0.9675**	0.9182	0.9361	0.9378	0.9603	**0.9711**	**0.9768**	**0.9778**
KRCC	0.7266	0.7605	**0.8424**	0.7467	0.7744	0.7793	0.8249	**0.8511**	**0.8677**	**0.8708**
PCC	0.8902	0.9159	**0.9646**	0.9030	0.9254	0.9244	0.9522	**0.9662**	**0.9762**	**0.9773**
RMSE	17.748	15.637	**10.268**	16.735	14.764	14.858	11.904	**10.048**	**8.4550**	**8.2480**
	Overall direct									
SRCC	0.8856	0.8750	0.8989	0.8723	0.9019	0.9048	**0.9385**	**0.9311**	**0.9252**	**0.9258**
KRCC	0.7084	0.6983	0.7442	0.6945	0.7300	0.7343	**0.7920**	**0.7897**	**0.7851**	**0.7865**
PCC	0.8767	0.8756	0.9021	0.8734	0.8965	0.9006	**0.9385**	**0.9349**	**0.9292**	**0.9295**
RMSE	0.6703	0.6830	0.6909	0.6787	0.6587	0.6417	**0.5178**	**0.5552**	**0.5957**	**0.5960**
	Overall weighted									
SRCC	0.8838	0.8649	0.8662	0.8633	0.8861	0.8875	**0.9263**	**0.9056**	**0.8945**	**0.8949**
KRCC	0.7048	0.6848	0.7011	0.6830	0.7078	0.7096	**0.7719**	**0.7515**	**0.7394**	**0.7400**
PCC	0.8741	0.8678	0.8718	0.8660	0.8817	0.8861	**0.9288**	**0.9137**	**0.9016**	**0.9016**
RMSE	0.7819	0.8068	0.8429	0.8068	0.7869	0.7708	**0.6211**	**0.6917**	**0.7442**	**0.7448**

In the table, pairwise score differences were used

he four best IQA measures for each evaluation index are written in bold. Overall results for RMSE do not take into account LIVE benchmark

respectively. The test revealed that lrSIMs are statistically better than other IQA measures, and in many cases, hybrid measures trained with PSD are better than hybrid measures trained with raw scores. Figure 2 presents summary of significance tests covering all benchmarks, including tests with PSD. For each benchmark, the numbers in cells were added. Since there are eight benchmarks, the number in the cell denotes the number of benchmarks in which the measure in the row is significantly better, or worse in case of the negative value,
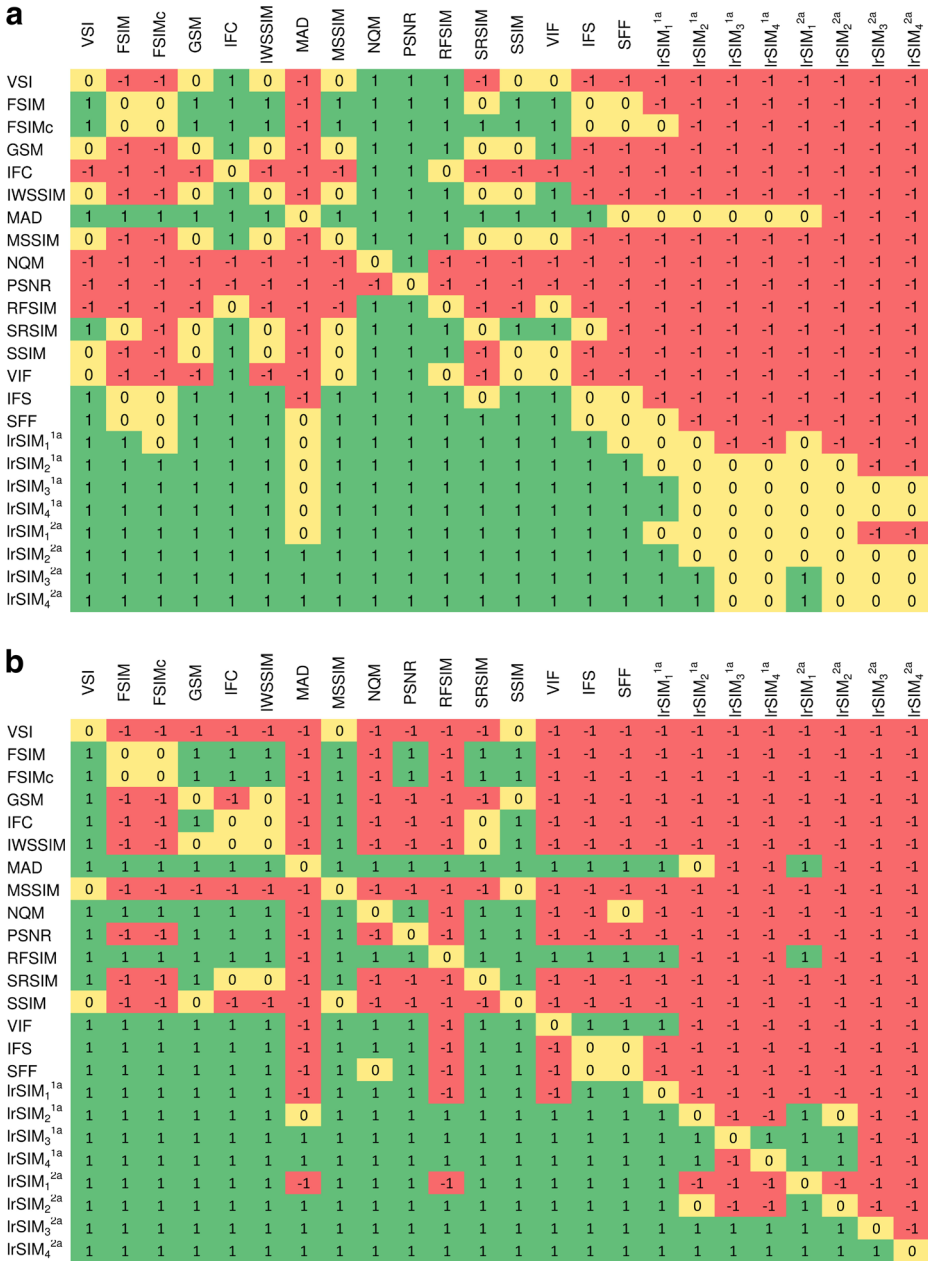
a

	VSI	FSIM	FSIMc	GSM	IFC	IWSSIM	MAD	MSSIM	NQM	PSNR	RFSIM	SRSIM	SSIM	VIF	IFS	SFF	$lrSIM_1^{1a}$	$lrSIM_2^{1a}$	$lrSIM_3^{1a}$	$lrSIM_4^{1a}$	$lrSIM_1^{2a}$	$lrSIM_2^{2a}$	$lrSIM_3^{2a}$	$lrSIM_4^{2a}$
VSI	0	-1	-1	0	1	0	-1	0	1	1	1	-1	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
FSIM	1	0	0	1	1	1	-1	1	1	1	1	0	1	1	0	0	-1	-1	-1	-1	-1	-1	-1	-1
FSIMc	1	0	0	1	1	1	-1	1	1	1	1	1	1	1	0	0	0	-1	-1	-1	-1	-1	-1	-1
GSM	0	-1	-1	0	1	0	-1	0	1	1	1	0	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
IFC	-1	-1	-1	-1	0	-1	-1	-1	1	1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
IWSSIM	0	-1	-1	0	1	0	-1	0	1	1	1	0	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
MAD	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	-1	-1	-1	-1
MSSIM	0	-1	-1	0	1	0	-1	0	1	1	1	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
NQM	-1	-1	-1	-1	-1	-1	-1	-1	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
PSNR	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
RFSIM	-1	-1	-1	0	-1	-1	-1	1	1	1	0	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
SRSIM	1	0	-1	0	1	0	-1	0	1	1	1	0	1	1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1
SSIM	0	-1	-1	0	1	0	-1	0	1	1	1	-1	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
VIF	0	-1	-1	-1	1	-1	-1	0	1	1	0	-1	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
IFS	1	0	0	1	1	1	-1	1	1	1	1	0	1	1	0	0	-1	-1	-1	-1	-1	-1	-1	-1
SFF	1	0	0	1	1	1	0	1	1	1	1	1	1	1	0	0	0	-1	-1	-1	-1	-1	-1	-1
$lrSIM_1^{1a}$	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	0	0	0	0	1	-1	-1	-1	-1
$lrSIM_2^{1a}$	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	-1	-1
$lrSIM_3^{1a}$	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
$lrSIM_4^{1a}$	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
$lrSIM_1^{2a}$	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	-1	-1
$lrSIM_2^{2a}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
$lrSIM_3^{2a}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0
$lrSIM_4^{2a}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0

b

	VSI	FSIM	FSIMc	GSM	IFC	IWSSIM	MAD	MSSIM	NQM	PSNR	RFSIM	SRSIM	SSIM	VIF	IFS	SFF	$lrSIM_1^{1a}$	$lrSIM_2^{1a}$	$lrSIM_3^{1a}$	$lrSIM_4^{1a}$	$lrSIM_1^{2a}$	$lrSIM_2^{2a}$	$lrSIM_3^{2a}$	$lrSIM_4^{2a}$
VSI	0	-1	-1	-1	-1	-1	-1	0	-1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
FSIM	1	0	0	1	1	1	-1	1	-1	1	-1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
FSIMc	1	0	0	1	1	1	-1	1	-1	1	-1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
GSM	1	-1	-1	0	-1	0	-1	1	-1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
IFC	1	-1	-1	1	0	0	-1	1	-1	-1	-1	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
IWSSIM	1	-1	-1	0	0	0	-1	1	-1	-1	-1	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
MAD	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	-1	-1	1	-1	-1	-1	-1
MSSIM	0	-1	-1	-1	-1	-1	-1	0	-1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
NQM	1	1	1	1	1	1	-1	1	0	1	-1	1	1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1
PSNR	1	-1	-1	1	1	1	-1	1	-1	0	-1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
RFSIM	1	1	1	1	1	1	-1	1	1	1	0	1	1	1	1	1	-1	-1	-1	1	-1	-1	-1	-1
SRSIM	1	-1	-1	0	0	0	-1	1	-1	-1	-1	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
SSIM	0	-1	-1	0	-1	-1	-1	0	-1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
VIF	1	1	1	1	1	1	-1	1	1	1	-1	1	1	0	1	1	-1	-1	-1	-1	-1	-1	-1	-1
IFS	1	1	1	1	1	1	-1	1	1	1	-1	1	1	-1	0	0	-1	-1	-1	-1	-1	-1	-1	-1
SFF	1	1	1	1	1	1	-1	1	0	1	-1	1	1	-1	0	0	-1	-1	-1	-1	-1	-1	-1	-1
$lrSIM_1^{1a}$	1	1	1	1	1	1	-1	1	1	1	-1	1	1	-1	1	1	0	-1	-1	-1	-1	-1	-1	-1
$lrSIM_2^{1a}$	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	-1	1	0	-1	-1	-1
$lrSIM_3^{1a}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	-1	-1	-1
$lrSIM_4^{1a}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-1	0	1	1	-1	-1	-1
$lrSIM_1^{2a}$	1	1	1	1	1	1	-1	1	1	1	-1	1	1	1	1	1	-1	-1	-1	0	-1	-1	-1	-1
$lrSIM_2^{2a}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	-1	-1	1	0	-1	-1	-1
$lrSIM_3^{2a}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	-1	
$lrSIM_4^{2a}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0

Fig. 1 Significance tests on LIVE benchmark using raw scores (**a**) and PSD (**b**). The number ''1'', ''0'' or ''-1'' in the *cell* denotes that the measure in the row is statistically better with the confidence greater than 95 %, indistinguishable, or worse than the measure in the column, respectively. It can be seen that lrSIMs are statistically better than state-of-the-art IQA measures on this dataset

	VSI	FSIM	FSIMc	GSM	IFC	IWSSIM	MAD	MSSIM	NQM	PSNR	RFSIM	SRSIM	SSIM	VIF	IFS	SFF	$lrSIM_1^{1a}$	$lrSIM_2^{1a}$	$lrSIM_3^{1a}$	$lrSIM_4^{1a}$	$lrSIM_1^{2a}$	$lrSIM_2^{2a}$	$lrSIM_3^{2a}$	$lrSIM_4^{2a}$
VSI	0	1	0	5	6	3	-2	6	6	6	1	0	6	2	-3	-3	-7	-4	-4	-4	-7	-4	-4	-4
FSIM	-1	0	-3	8	8	6	0	7	6	8	0	-1	8	2	-5	-6	-8	-5	-5	-7	-8	-7	-5	-5
FSIMc	0	3	0	8	8	7	0	8	6	8	0	4	8	3	-2	-3	-7	-4	-4	-4	-8	-7	-4	-4
GSM	-5	-8	-8	0	6	-3	-3	2	6	6	-2	-7	6	2	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
IFC	-6	-8	-8	-6	0	-7	-8	-6	-1	3	-7	-7	-6	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
IWSSIM	-3	-6	-7	3	7	0	-1	5	6	6	-1	-4	7	2	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
MAD	2	0	0	3	8	1	0	5	8	8	3	0	8	8	-2	-3	-4	-6	-7	-7	-5	-8	-8	-8
MSSIM	-6	-7	-8	-2	6	-5	-5	0	6	6	-3	-7	6	1	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
NQM	-6	-6	-6	-6	1	-6	-8	-6	0	3	-8	-6	-6	-8	-8	-7	-8	-8	-8	-8	-8	-8	-8	-8
PSNR	-6	-8	-8	-6	-3	-6	-8	-6	-3	0	-8	-6	-6	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
RFSIM	-1	0	0	2	7	1	-3	3	8	8	0	-1	6	4	-3	-5	-6	-8	-8	-8	-6	-8	-8	-8
SRSIM	0	1	-4	7	7	4	0	7	6	6	1	0	8	3	-4	-7	-8	-5	-6	-6	-8	-7	-6	-6
SSIM	-6	-8	-8	-6	6	-7	-8	-6	6	6	-6	-8	0	-4	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
VIF	-2	-2	-3	-2	8	-2	-8	-1	8	8	-4	-3	4	0	-6	-6	-6	-8	-8	-8	-8	-8	-8	-8
IFS	3	5	2	8	8	8	2	8	8	8	3	4	8	6	0	-4	-7	-4	-4	-3	-8	-6	-4	-3
SFF	3	6	3	8	8	8	3	8	7	8	5	7	8	6	4	0	-6	-5	-5	-3	-7	-7	-5	-4
$lrSIM_1^{1a}$	7	8	7	8	8	8	4	8	8	8	6	8	8	6	7	6	0	-2	-1	1	-4	-3	0	0
$lrSIM_2^{1a}$	4	5	4	8	8	8	6	8	8	8	8	5	8	8	4	5	2	0	-2	2	2	-2	-2	-2
$lrSIM_3^{1a}$	4	5	4	8	8	8	7	8	8	8	8	6	8	8	4	5	1	2	0	3	-1	-1	-2	-1
$lrSIM_4^{1a}$	4	7	4	8	8	8	7	8	8	8	8	6	8	8	3	3	-1	-2	-3	0	-4	-4	-2	-2
$lrSIM_1^{2a}$	7	8	8	8	8	8	5	8	8	8	6	8	8	8	8	7	4	-2	1	4	0	-2	-1	-1
$lrSIM_2^{2a}$	4	7	7	8	8	8	8	8	8	8	8	7	8	8	6	7	3	2	1	4	2	0	1	1
$lrSIM_3^{2a}$	4	5	4	8	8	8	8	8	8	8	8	6	8	8	4	5	0	2	2	2	1	-1	0	-2
$lrSIM_4^{2a}$	4	5	4	8	8	8	8	8	8	8	8	6	8	8	3	4	0	2	1	2	1	-1	2	0

Fig. 2 Summary of significance tests on four benchmarks, including tests with PSD (eight benchmarks in total). For each test, the number "1", "0" or "-1" in the *cell* denotes that the measure in the row is statistically better with the confidence greater than 95 %, indistinguishable, or worse than the measure in the column, respectively. The values in cells for eight tests were added in order to form this summary. The higher value in the cell indicates the better IQA measure in the row

than the measure in the column. Taking into account the summary of significance tests, it can be seen that $lrSIM_2^{2a}$ is the best performing measure, with non-negative values in cells. It is worth noticing that $lrSIM_2^{2a}$ is worse that $lrSIM_1^{2a}$ if only significance tests with non-hybrid measures are taken into account. All lrSIMs have non-negative values in cells shared with non-hybrid measures, where they are in rows, and hybrid IQA measures developed with PSD, $lrSIM_{1-4}^{2a}$, are at the top of the ranking. Among non-hybrid approaches, SFF is the leading IQA measure, followed by FSIMc, MAD and VSI.

The experimental evaluation showed that it is worth using PSD in training of the proposed hybrid IQA measure family. This can also be seen on scatter plots with subjective opinion scores against objective scores of the two best IQA measures and $lrSIM_1^{2a}$ on four benchmarks (see Fig. 3). Here, $lrSIM_1^{2a}$ is better correlated with subjective scores than compared measures.

Since in this paper the hybrid approach is presented, it should be compared with other similar approaches that are present in the literature. Therefore, Table 5 contains comparison with such approaches on the basis of published SRCC values. This also gives the opportunity to compare the results with non-hybrid IQA measures, which also can be found in the literature. The table contains results for TID2008, CSIQ and LIVE benchmarks, since most
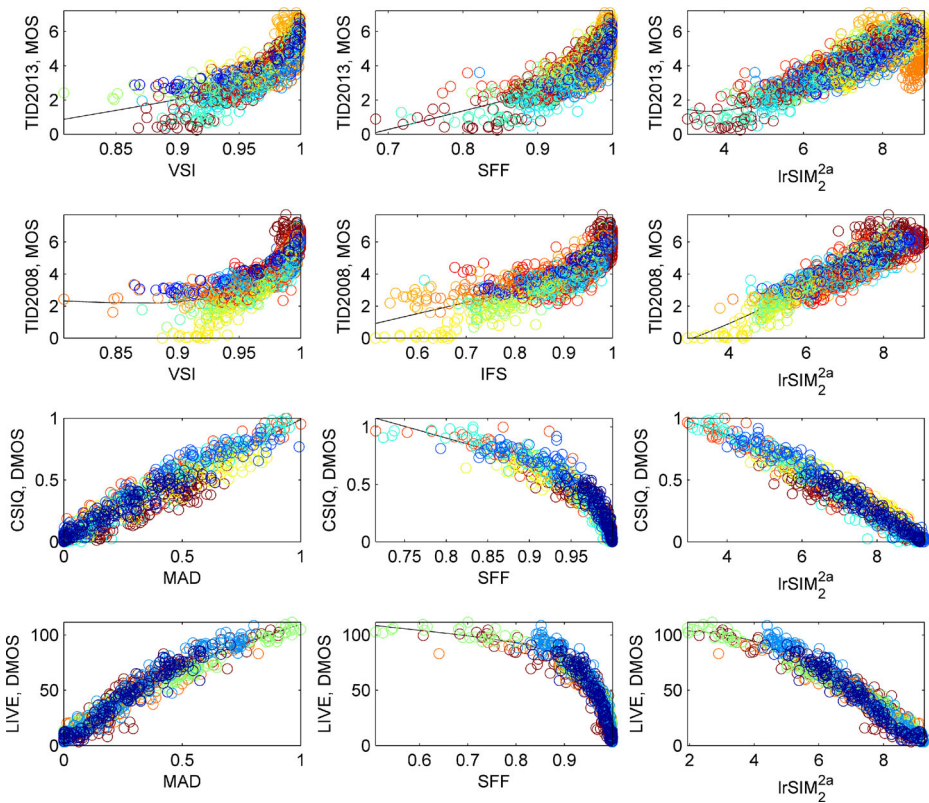
Fig. 3 Scatter plots of subjective opinion scores against subjective scores of the two best IQA measures and $lrSIM_1^{2a}$ on four benchmarks. Plots also contain *curves* fitted with logistic functions, names of benchmark datasets (*vertical axis*) and IQA measures (*horizontal axis*). *Colours* represent different distortions; each dataset has its own set of colours

of the compared measures were not evaluated on TID2013. Here, the best three results for a given benchmark are written in boldface, results not reported are denoted by "-". Furthermore, "-" denotes overall results for IQA measures that were not evaluated on these three benchmarks or are not independent, i.e., authors developed a separate measure for each benchmark without providing cross-benchmark results ([13, 18, 20, 22, 25, 48]).

The results of comparison based on SRCC are presented in Table 5. They reveal that lrSIMs outperformed other measures, being in most cases among the three best IQA techniques. For TID2008, SM-HM-FSIM [11] was the second best technique. However, SM-HM-FSIM is non-hybrid approach, what makes all lrSIMs better than other compared hybrid approaches. The presented lrSIM family, together with the approach introduced by Barri et al. [2] outperformed other techniques on CSIQ. Overall results, as well as tests on LIVE, showed superior performance of lrSIMs over the other measures. More specifically, $lrSIM_1^{2a}$ and $lrSIM_2^{2a}$ were clearly better than other measures, what was indicated in the previous experiments.

Table 5 Comparison of the approach with other IQA approaches and hybrid measures based on SRCC values reported in the literature

IQA measure	TID2008	CSIQ	LIVE	Overall direct	Overall weighted
	Other IQA measures				
ESSIM [53]	*0.8843*	0.9326	0.9622	0.9264	0.9151
GMSD [45]	0.8910	0.9570	0.9600	0.9360	0.9244
SURF-SIM [38]	0.8910	0.9460	0.9440	0.9270	0.9178
[44]	*0.9080*	*0.9650*	*0.9640*	–	–
[54]	*0.8952*	*0.9640*	*0.9737*	–	–
SM-HM-FSIM [11]	**0.8960**	0.9530	0.9640	0.9377	0.9241
	Other hybrid measures				
ADM [15]	0.8617	0.9333	0.9460	0.9137	0.9001
[2]	0.8100	**0.9630**	*0.9570*	0.9100	0.8843
BME [25]	*0.8882*	*0.9573*	*0.9711*	–	–
BMMF [12]	*0.9471*	–	–	–	–
CNNM [20]	–	–	–	–	–
CQM [21]	*0.8720*	–	–	–	–
DOG-SSIM [24]	*0.9259*	0.9204	0.9423	0.9295	0.9282
EHIS [22]	*0.9098*	*0.9498*	*0.9622*	–	–
ESIM [23]	*0.9026*	0.9620	0.9420	0.9420	**0.9300**
GLD-PFT [30]	0.8849	0.9549	0.9631	0.9343	0.9186
IGM [43]	0.8902	0.9401	0.9580	0.9294	0.9190
[13]	–	–	*0.9500*	–	–
MAD [14]	0.8340	0.9466	0.9669	0.9158	0.8944
MMF [18]	*0.9487*	*0.9755*	*0.9732*	–	–
RMSSIM [26]	0.8569	*0.9453*	*0.9633*	–	–
[48]	0.8865	*0.9141*	*0.9574*	–	–
lrSIM$_1^{2a}$	**0.9013**	**0.9630**	**0.9691**	**0.9445**	**0.9309**
lrSIM$_2^{2a}$	*0.9073*	**0.9688**	**0.9730**	**0.9497**	**0.9365**
lrSIM$_3^{2a}$	0.8916	*0.9670*	**0.9732**	0.9439	0.9275
lrSIM$_4^{2a}$	**0.8923**	**0.9681**	*0.9735*	**0.9447**	0.9283

The result for the measure that was trained using images from the dataset indicated in the column is italicised in order to show the lack of the dataset-independence. Overall results exclude IQA measures for which cross-dataset results were not reported in the literature. The three best measures for each benchmark are shown in boldface

4 Conclusions

In this paper, a hybrid full-reference IQA was introduced. The measure was obtained using the lasso regression and pairwise score differences of up to 16 IQA measures seen as predictors. The lasso was able to select the most important several IQA measures. This resulted

in the family of hybrid measures, lrSIMs, which was extensively evaluated on four largest IQA image benchmarks employing SRCC, KRCC, PCC, and RMSE. The evaluation was also based on PSD. The introduced approach outperformed widely used full reference IQA measures, as well as other hybrid techniques. It can be assumed that the usage of PSD will support the development of other IQA measures based on supervised learning.

The Matlab code of the approach that allows adding other IQA measures, scripts performing pairwise score differences on used benchmarks, as well as the evaluation of compared approaches, are available to download at: http://marosz.kia.prz.edu.pl/lrSIMpsd.html.

References

1. Attar A, Shahbahrami A, Rad RM (2015) Image quality assessment using edge based features. Multimed Tools Appl:1–16. doi:10.1007/s11042-015-2663-9
2. Barri A, Dooms A, Jansen B, Schelkens P (2014) A locally adaptive system for the fusion of objective quality measures. IEEE T Image Process 23(6):2446–2458. doi:10.1109/TIP.2014.2316379
3. Bhandari AK, Kumar A, Chaudhary S, Singh GK (2015) A new beta differential evolution algorithm for edge preserved colored satellite image enhancement. Multidim Syst Sign:1–33. doi:10.1007/s11045-015-0353-4
4. Chandler DM (2013) Seven challenges in image quality assessment: Past, present, and future research. ISRN Signal Process 2013:53. doi:10.1155/2013/905685
5. Chandra A, Chattopadhyay S (2014) A new strategy of image denoising using multiplier-less fir filter designed with the aid of differential evolution algorithm. Multimed Tools Appl 75(2):1079–1098. doi:10.1007/s11042-014-2358-7
6. Chang HW, Yang H, Gan Y, Wang MH (2013) Sparse feature fidelity for perceptual image quality assessment. IEEE T Image Process 22(10):4007–4018. doi:10.1109/TIP.2013.2266579
7. Chang HW, Zhang QW, Wu QQ, Gan Y (2015) Perceptual image quality assessment by independent feature detector. Neurocomputing 151(part 3):1142–1152. doi:10.1016/j.neucom.2014.04.081
8. Damera-Venkata N, Kite TD, Geisler WS, Evans BL, Bovik AC (2000) Image quality assessment based on a degradation model. IEEE T Image Process 9(4):636–650. doi:10.1109/83.841940
9. Gao X, Lu W, Tao D, Li X (2010) Image quality assessment and human visual system. In: Visual communications and image processing 2010, international society for optics and photonics, pp 77,440Z–77,440Z
10. Group VQE (2003, [Accessed 26.06.2016]) Final report from the video quality experts group on the validation of objective models of video quality assessment, phase ii (fr_tv2). https://wwwituint/ITU-T/studygroups/com09/docs/tutorial_opavcpdf
11. Huang Y, Chen X, Ding X (2015) A harmonic means pooling strategy for structural similarity index measurement in image quality assessment. Multimed Tools Appl:1–12. doi:10.1007/s11042-015-2620-7
12. Jin L, Egiazarian K, Kuo CC (2012) Perceptual image quality assessment using block-based multi-metric fusion (BMMF). In: Proc. IEEE Int. conf. on acoustics, speech and signal processing (ICASSP), pp 1145–1148. doi:10.1109/ICASSP.2012.6288089
13. Lahouhou A, Viennet E, Beghdadi A (2010) Selecting low-level features for image quality assessment by statistical methods. CIT 18(2)
14. Larson EC, Chandler DM (2010) Most apparent distortion: full-reference image quality assessment and the role of strategy. J Electron Imaging 19(1):011006. doi:10.1117/1.3267105
15. Li S, Zhang F, Ma L, Ngan KN (2011) Image quality assessment by separately evaluating detail losses and additive impairments. IEEE T Multimed 13(5):935–949. doi:10.1109/TMM.2011.2152382
16. Liu A, Lin W, Narwaria M (2012) Image quality assessment based on gradient similarity. IEEE T Image Process 21(4):1500–1512. doi:10.1109/tip.2011.2175935

17. Liu M, Yang X (2009) A new image quality approach based on decision fusion. In: Ma J, Yin Y, Yu J, Zhou S (eds) Proc. Int. conf. on fuzzy systems and knowledge discovery (FSKD). IEEE, pp 10–14

18. Liu TJ, Lin W, Kuo CC (2013) Image quality assessment using multi-method fusion. IEEE T Image Process 22(5):1793–1807. doi:10.1109/TIP.2012.2236343

19. Lu W, Mei N, Gao F, He L, Gao X (2015) Blind image quality assessment via semi-supervised learning and fuzzy inference. Applied Informatics 2(1):1–20. doi:10.1186/s40535-015-0010-x

20. Lukin VV, Ponomarenko NN, Ieremeiev OI, Egiazarian KO, Astola J (2015) Combining full-reference image visual quality metrics by neural network. In: Human vision and electronic imaging XX, Proc. SPIE, vol 9394, p 93940K. doi:10.1117/12.2085465

21. Okarma K (2010) Combined full-reference image quality metric linearly correlated with subjective assessment. In: Artificial intelligence and soft computing. Springer, pp 539–546. doi:10.1007/978-3-642-13208-7_67

22. Okarma K (2013) Extended Hybrid Image Similarity - combined full-reference image quality metric linearly correlated with subjective scores. Elektronika ir Elektrotechnika 19(10):129–132

23. Oszust M (2016) Decision fusion for image quality assessment using an optimization approach. IEEE Signal Proc Let 23(1):65–69. doi:10.1109/LSP.2015.2500819

24. Pei SC, Chen LH (2015) Image quality assessment using human visual dog model fused with random forest. IEEE T Image Process 24(11):3282–3292. doi:10.1109/TIP.2015.2440172

25. Peng P, Li ZN (2012) A mixture of experts approach to multi-strategy image quality assessment. In: Campilho A, Kamel M (eds) Image analysis and recognition, lecture notes in computer science, vol 7324. Springer, Berlin-Heidelberg, pp 123–130, doi:10.1007/978-3-642-31295-3_15

26. Peng P, Li ZN (2012) Regularization of the structural similarity index based on preservation of edge direction. In: 2012 IEEE International conference on systems, man, and cybernetics (SMC), pp 2127–2132. doi:10.1109/ICSMC.2012.6378054

27. Ponomarenko N, Lukin V, Zelensky A, Egiazarian K, Carli M, Battisti F (2009) TID2008 - a database for evaluation of full-reference visual quality assessment metrics. Adv Modern Radioelectron 10:30–45

28. Ponomarenko N, Jin L, Ieremeiev O, Lukin V, Egiazarian K, Astola J, Vozel B, Chehdi K, Carli M, Battisti F, Kuo CCJ (2015) Image database TID2013: peculiarities results and perspectives. Signal Process-Image 30:57–77. doi:10.1016/j.image.2014.10.009

29. Rehman A, Rostami M, Wang Z, Brunet D, Vrscay ER (2012) Ssim-inspired image restoration using sparse representation. EURASIP J Adv Sign Process 2012(1):1–12. doi:10.1186/1687-6180-2012-16

30. Saha A, Wu QMJ (2014) Full-reference image quality assessment by combining global and local distortion measures. CoRR abs/1412.5488

31. Shahid M, Rossholm A, Lövström B, Zepernick HJ (2014) No-reference image and video quality assessment: a classification and review of recent approaches. EURASIP Int J Image Video Process 2014(1):1–32. doi:10.1186/1687-5281-2014-40

32. Sharif M, Hussain A, Jaffar MA, Choi TS (2014) Fuzzy similarity based non local means filter for rician noise removal. Multimed Tools Appl 74(15):5533–5556. doi:10.1007/s11042-014-1867-8

33. Sheikh H, Bovik A, de Veciana G (2005) An information fidelity criterion for image quality assessment using natural scene statistics. IEEE T Image Process 14(12):2117–2128. doi:10.1109/tip.2005.859389

34. Sheikh HR, Bovik AC (2006) Image information and visual quality. IEEE T Image Process 15(2):430–444. doi:10.1109/TIP.2005.859378

35. Sheikh HR, Sabir MF, Bovik AC (2006) A statistical evaluation of recent full reference image quality assessment algorithms. IEEE T Image Process 15(11):3440–3451. doi:10.1109/tip.2006.881959

36. Tibshirani R (1996) Regression shrinkage and selection via the lasso. J R Stat Soc:267–288

37. Tibshirani R (2011) Regression shrinkage and selection via the lasso: a retrospective. J R Stat Soc 73(3):273–282

38. Wang F, Sun X, Guo Z, Huang Y, Fu K (2015) An object-distortion based image quality similarity. IEEE Signal Proc Let IEEE 22(10):1534–1537. doi:10.1109/LSP.2015.2413891

39. Wang Z, Bovik AC (2002) A universal image quality index. IEEE Signal Proc Let 9(3):81–84. doi:10.1109/97.995823

40. Wang Z, Li Q (2011) Information content weighting for perceptual image quality assessment. IEEE T Image Process 20(5):1185–1198. doi:10.1109/tip.2010.2092435

41. Wang Z, Simoncelli EP, Bovik AC (2003) Multi-scale structural similarity for image quality assessment. In: Proc. IEEE Int. conf. on signals, Systems, and Computers (ASILOMAR), pp 1398–1402

42. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE T Image Process 13(4):600–612. doi:10.1109/tip.2003.819861

43. Wu J, Lin W, Shi G, Liu A (2013) Perceptual quality metric with internal generative mechanism. IEEE T Image Process 22(1):43–54. doi:10.1109/TIP.2012.2214048

44. Wu J, Lin W, Shi G (2014) Image quality assessment with degradation on spatial structure. IEEE Signal Proc Let IEEE 21(4):437–440. doi:10.1109/LSP.2014.2304714
45. Xue W, Zhang L, Mou X, Bovik AC (2014) Gradient magnitude similarity deviation: a highly efficient perceptual image quality index. IEEE T Image Process 23(2):684–695. doi:10.1109/tip.2013.2293423
46. Yang J, Hou C, Xu R, Lei J (2010) New metric for stereo image quality assessment based on hvs. Int J Imaging Syst Technol 20(4):301–307. doi:10.1002/ima.20246
47. Yang J, Lin Y, Gao Z, Lv Z, Wei W, Song H (2015) Quality index for stereoscopic images by separately evaluating adding and subtracting. PLoS ONE 10(12):1–19. doi:10.1371/journal.pone.0145800
48. Yuan Y, Guo Q, Lu X (2015) Image quality assessment: a sparse learning way. Neurocomputing 159:227–241. doi:10.1016/j.neucom.2015.01.066
49. Zhang L, Li H (2012) SR-SIM: a fast and high performance IQA index based on spectral residual. In: Proc. IEEE Int. conf. on image processing (ICIP). IEEE, doi:10.1109/icip.2012.6467149
50. Zhang L, Zhang L, Mou X (2010) RFSIM: a feature based image quality assessment metric using riesz transforms. In: Proc. IEEE Int. conf. on image processing (ICIP). IEEE, doi:10.1109/icip.2010.5649275
51. Zhang L, Zhang L, Mou X, Zhang D (2011) FSIM: A feature similarity index for image quality assessment. IEEE T Image Process 20(8):2378–2386. doi:10.1109/tip.2011.2109730
52. Zhang L, Shen Y, Li H (2014) VSI: a visual saliency-induced index for perceptual image quality assessment. IEEE T Image Process 23(10):4270–4281. doi:10.1109/tip.2014.2346028
53. Zhang X, Feng X, Wang W, Xue W (2013) Edge strength similarity for image quality assessment. IEEE Signal Proc Let 20(4):319–322. doi:10.1109/LSP.2013.2244081
54. Zhou F, Lu Z, Wang C, Sun W, Xia ST, Liao Q (2015) Image quality assessment based on inter-patch and intra-patch similarity. PLoS ONE 10(3):e0116,312. doi:10.1371/journal.pone.0116312

Mariusz Oszust received the MSc degree in electrical engineering from the Rzeszow University of Technology in 2005 and PhD in computer science in 2013 from AGH University of Science and Technology, Krakow, Poland. Currently he is an assistant professor at the Rzeszow University of Technology. His main research interests include pattern recognition, optimisation, image quality assessment and development of vision-based human-computer interfaces.