SHEARLET-BASED REDUCED REFERENCE IMAGE QUALITY ASSESSMENT

Sebastian Bosse[1], *Qiaobo Chen*[1,2], *Mischa Siekmann*[1],
Wojciech Samek[1], Member, IEEE, and *Thomas Wiegand*[1,2], Fellow, IEEE

[1] Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, Berlin, Germany
[2] Department of Electrical Engineering, Technical University of Berlin, Germany.

ABSTRACT

This paper proposes a reduced reference image quality assessment method using only a low number of features. It involves a shearlet decomposition, directional pooling of the obtained coefficient and extracts the scalewise statistical location parameter as a feature. The proposed method is tested and compared to similar approaches on the LIVE image database. On this database it outperforms the compared methods on five of seven distortion types and on the full testset with a linear correlation of $r_{all} = 0.89$.

***Index Terms*—** Shearlets, Inverse Gaussian distribution, natural image statistics, image quality assessment, reduced-reference

1. INTRODUCTION

Visual information plays an important role in everyday life. Almost 50% of the human brain is involved in visual processing and 70% of all human sensory receptors are optical [1]. In recent years the availability of visual media has increased in numerous applications, most of which are meant for the human visual system as the ultimate receiver. However, due to the technical necessity to compress images and videos in most applications (e.g. for the sake of bandwidth saving), degradations occur that are visible to humans. Thus, perceptual image quality assessment (IQA) is a crucial element for the optimization of image and video communication systems. Generally, IQA approaches can be categorized according to the amount of information about the undistorted reference image used for estimating the perceived quality of the distorted image. While full reference (FR) approaches have the whole reference image available, no reference (NR) do not make use of any specific information about the reference image. Popular examples of FR-IQA methods are [2, 3, 4, 5, 6], in the relatively new field of NR-IQA [7, 8, 9, 10, 11, 12] are promising approaches. Within this range, reduced reference (RR) IQA is living somewhere in the middle, as only a set of features extracted from reference image is accessible by the RR-IQA method.

In this paper we propose a novel approach to RR-IQA based on a shearlet decomposition and suggest an appropriate distribution model for the resulting coefficients. The performance of the proposed RR-IQA approach is evaluated by comparison to similar approaches based on the LIVE database [13].

2. REDUCED REFERENCE IMAGE QUALITY ASSESSMENT

Algorithms following the RR-IQA paradigm extract quality characterizing features from the reference image at the sender side that are transferred to receiver side. At receiver side, the corresponding features are extracted from the transmitted and potentially distorted image and compared to those transmitted reference features. The difference between the two sets of features is then used for estimating the perceived quality of the transmitted image.

[14] presents a framework that makes use of a channel decomposition based on the steerable pyramid [15]. The Kullback-Leibler divergence between the marginal probability distributions of the coefficients of the reference and distorted images in different subbands is used as an estimate of perceived image quality. The marginal distributions of the reference image coefficients are summarized by the parameters of a generalized Gaussian model. [16] follows a similar approach, but normalizes the wavelet coefficients using a divisive normalization transform. By this, the coefficients can be modeled using a zero-mean Gaussian model. In [17], coefficients are obtained by a grouplet transform and thresholded according to contrast sensitivity and masking. A similar approach where features are extracted in a perceptual space is presented in [18]. A family of information theoretic RR-IQA algorithms is developed in [19]. For this, coefficients obtained by scale-space-orientation decomposition (implemented as a steerable pyramid) are modeled using Gaussian scale mixture distributions and it is studied how the performance decays with reduction or distortion of the information taken from the reference image. Based on the observation that image gradient magnitude follows a Weibull distribution, in [20] the absolute values of the coefficients of a steerable pyramid are directionally pooled for each scale and modeled by a Weibull distribution. The scale parameters of the Weibull distribution is scalewise used as a RR-feature. The presented work follows a similar line.

3. METHODS

3.1. Shearlets

Recently, a directional multiscale decomposition system known as shearlets has been developed, which provides a unified treatment of continuous as well as discrete models, allowing optimally sparse representations of piecewise smooth images [21, 22]. Shearlets are constructed by anisotropically scaling, shearing and translating a (wavelet analogous) shearlet generator $\phi(t)$. In order to preserve the integer lattice of the image, the change of orientation is not achieved by rotation, but by shearing the system and thus permitting faithful digital implementations by

$$\mathbf{O_s} = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \tag{1}$$

This is the main and distinctive difference to curvelets and its discrete form contourlets, which control directional selectivity by rotation [23]. Scaling is performed using a scaling matrix

$$\mathbf{A}_a = \begin{bmatrix} a & 0 \\ 0 & a^\alpha \end{bmatrix}, \tag{2}$$

where $\alpha \in [0, 1]$ determines the degree of anisotropy. Finally, a translation operator shifts the generating function $\mathbf{T}_m\phi(t) \rightarrow \phi(t - m)$, yielding a shearlet system $S(\phi) = \mathbf{T}_m \cdot \mathbf{A}_a \cdot \mathbf{O}_s \cdot \phi$.

For a certain class of natural images, so-called cartoon-like images (images consisting of smooth areas separated by piecewise smooth boundary curves), shearlet-based transforms were shown to provide optimally sparse approximations. This simplified model of natural images emphasizing anisotropic features, most notably edges, is found to be consistent with many models of the human visual system [23].

3.2. Inverse Gaussian Distribution

The inverse Gaussian distribution (also know as Wald-distribution) is a two parameter probability distribution with a position parameter $\mu > 0$ and a shape parameter $\lambda > 0$. The probability density function is defined as

$$f_{IG}(x; \mu, \lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} \cdot e^{\frac{-\lambda(x-\mu)^2}{2\mu^2 x}} \tag{3}$$

The position parameter μ in f_{IG} can easily be estimated as the mean of the data to be summarized.

Note that the inverse Gaussian distribution is very similar to the Weibull distribution that is assumed to summarize the gradient magnitude of natural images in a sequential fragmentation process [24]. The Weibull distribution was used in a similar context in [20]. However, in [25], normal inverse Gaussian distributions (a normal variance-mean mixture distribution with a mixing density of inverse Gaussian) have been shown to be a feasible model for curvelet coefficients. Given the conceptual similarity of curvelets and shearlets (see Subsec.3.1), this motivates the use of inverse Gaussian distribution parameters.

3.3. Scalewise orientational coefficient pooling

Similar to the method proposed in [20], for each scale n the absolute values of the shearlet coefficients $c_{n,k}(x, y)$ are max-pooled samplewise along the orientational shearings k. This leads to the scalewise feature map

$$c_n(x, y) = \max_{k \in [0,\ldots,K-1]} |c_{n,k}(x, y)| . \tag{4}$$

This kind of pooling is also neurophysiologically plausible for modeling visual processing in the human brain [26].

3.4. Distribution Fitting

For each frequency scale n the distributions position parameters μ_n are estimated as the mean value of scalewise orientational pooled feature map $f_n(x, y)$ with

$$\mu_n = \frac{1}{XY} \sum_{x=0}^{X} \sum_{y=0}^{Y} c_n(x, y) \tag{5}$$

and X, Y being the height and width of the specific image.

3.5. Feature Pooling

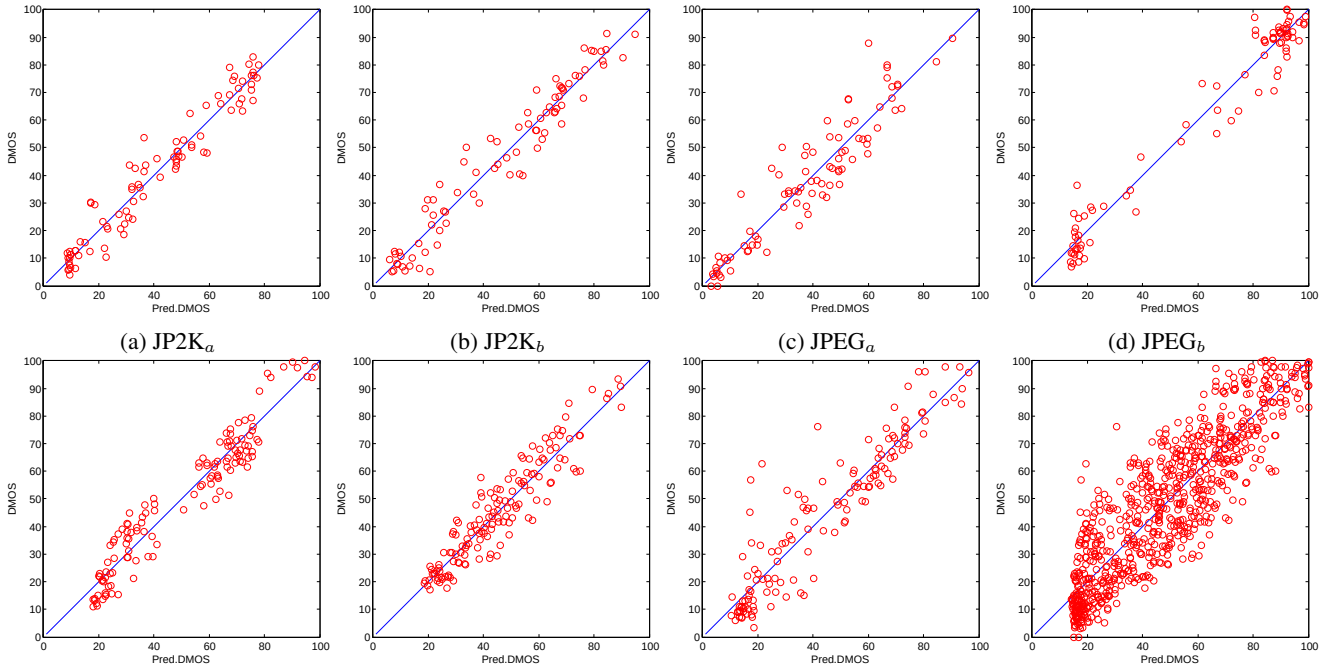
The resulting scalewise features are calculated for the reference image as μ_n^R and for the distorted image as μ_n^D. μ_n^R is an indicator of the activity in the reference image on a certain scale n. As such, it can be used to model global masking effects on a certain scale and control for the global distortion sensitivity of an image. Similar to [20], this leads to a RR quality metric

$$D = \sum_{n=0}^{N} \frac{\left|\mu_n^R - \mu_n^D\right|}{\mu_n^R} \tag{6}$$

4. EXPERIMENTS

In order to test the proposed RR IQA method and to be able to compare it to [20], we signed a shearlet system with $N = 5$ frequency scales (plus low-pass component) and the same number of $K = 8$ orientations on each scale. The shearlet system is based on a 9-tap quadrature mirror filter.

Evaluations and comparison were performed on the LIVE database [13]. It contains 29 color reference images and 779 distorted images, distorted by JPEG compression (JPEG), JPEG2000 compression (JP2K), additive Gaussian white noise (WN), Gaussian blurring (GB) and JPEG2000 compressed images transmitted over a simulated Rayleigh fading channel (FF). JPEG and JPEG2000 compressed images are organized in two sets. Each distortion type is introduced at



(a) JP2K$_a$ (b) JP2K$_b$ (c) JPEG$_a$ (d) JPEG$_b$

(e) White noise (f) Gaussian blur (g) Fast fading (h) All

Fig. 1: DMOS values from LIVE database scattered vs. predicted values from the proposed method for the 7 distortion categories and all distortions.

5-6 magnitudes. On average, about 23 subjects evaluated the quality of each image. The viewing conditions are fairly controlled in terms of viewing distance. Ratings were collected in a double stimulus manner and reported as Differential Mean Opinion Scores (DMOS). Although the LIVE database contains color images, for the evaluation images were converted to gray scale.

Following [27], a 5-parameter logistic function

$$f(x) = a_1 \left(\frac{1}{2} - \frac{1}{1 + e^{(a_2 \cdot (x - a_3))}} \right) + a_4 x + a_5 \quad (7)$$

was used to map the obtained RR quality metric D nonlinearily to the DMOS values from the LIVE database.

5. RESULTS

Table 1 shows the performance of the proposed RR IQA method (indicated by 'SL+invGauss-fit') in terms of prediction accuracy measured as Pearson product-moment correlation coefficient (CC) and prediction monotonicity measured by Spearman's rank order correlation coefficient (SROCC). The performance is compared to the method described in [20], that uses a decomposition based on a steerable pyramid and summarizes the resulting coefficients by the shape parameter of a Weibull distribution. In order to evaluate the influence of the type of the multiscale decomposition and of the type of the summarizing statistics, we further compare the method to a combination of the described shearlet based decomposition with the scaling parameter from the Weibull distribution ('SL+wbl-fit') and did the same recombination with the method from [20], by combining a steerable pyramid based decomposition with a position parameter from the inverse Gaussian distribution ('SP+invGauss-fit'). (Unfortunately, we were not able to reproduce the results from [20], so we report the correlations achieved with our implementation of a combination of steerable pyramid with a Weibull-scaling parameter (SP+wbl-fit) as well.) To bring the results into a broader context, the correlations for two FR IQA methods (PSNR and MS-SSIM [4]) are reported as well. In terms of CC and SROCC, the proposed method performs best for JP2000 and JPEG compression artifacts on both sets in the database and for fast fading distortions among the RR IQA methods. For additive white Gaussian noise and Gaussian blur type distortions, method [20] obtains a higher CC and for WN also a higher SROCC. Changing the assumption about the underlying distribution to a Weibull model for a shearlet-based decomposition achieves the highest SROCC for Gaussian blurring. Most interestingly, among the evaluated RR IQA methods the proposed method achieves the highest prediction accuracy and prediction monotonicity, indicating good generalization properties. In Fig.1 the performance of the proposed methods for the LIVE database is shown graphically by scatter plots.

		JP2K$_a$	JP2K$_b$	JPEG$_a$	JPEG$_b$	WN	GB	FF	all
	Pearson Correlation Coefficient (CC)								
RR	SL+invGauss-fit	**0.9697**	**0.9702**	**0.9366**	**0.9799**	0.9634	0.9311	**0.9391**	**0.8875**
RR	SL+wbl-fit	0.9436	0.9409	0.8048	0.9641	0.9439	0.9125	0.9390	0.8412
RR	Method from [20]	0.9514	0.9569	0.8673	0.9568	**0.9755**	**0.9454**	0.9243	0.8353
RR	SP+invGauss-fit	0.9643	0.9469	0.9008	0.8476	0.9400	0.6669	0.8764	0.7266
RR	SP+wbl-fit	0.9488	0.9136	0.8278	0.9159	0.9405	0.6387	0.8721	0.7163
FR	*PSNR*	0.9332	0.8740	0.8856	0.9167	0.9859	0.7834	0.8895	0.8709
FR	*MS-SSIM* [4]	0.9702	0.9711	0.9699	0.9879	0.9737	0.9487	0.9304	0.9393
	Spearman Rank Order Correlation Coefficient (SROCC)								
RR	SL+invGauss-fit	**0.9673**	**0.9662**	**0.9274**	**0.9084**	0.9607	0.9267	**0.9370**	**0.8840**
RR	SL+wbl-fit	0.9402	0.9434	0.7894	0.8646	0.9300	0.9076	0.9310	0.8412
RR	Method from [20]	0.9595	0.9517	0.8535	0.8705	**0.9715**	**0.9371**	0.9258	0.8391
RR	SP+invGauss-fit	0.9583	0.9391	0.9008	0.7737	0.9597	0.6823	0.8857	0.7327
RR	SP+wbl-fit	0.9461	0.9029	0.8170	0.8307	0.9559	0.6538	0.8497	0.7214
FR	*PSNR*	0.9263	0.8549	0.8779	0.7708	0.9854	0.7823	0.8907	0.8755
FR	*MS-SSIM* [4]	0.9645	0.9648	0.9702	0.9454	0.9805	0.9519	0.9396	0.9527

Table 1: Performance of the proposed RR-IQA method on the LIVE database categorized by different distortion type subsets of the database. Prediction accuracy and prediction monotonicity are compared for combinations of decompositions using shearlets (SL) and a steerable pyramid (SP) with position parameters of inverse Gaussian (invGauss-fit) and Weibull distributions (wbl-fit). Further, the correlations reported in [20] and those of the FR methods PSNR and MS-SSIM are listed. Highest correlations among RR methods are indicated by bold number.

6. DISCUSSION

In this article a RR IQA method based on the statistical characterization of pooled shearlet coefficients based on the location parameters of a inverse Gaussian distribution was proposed. The location parameter of the inverse Gaussian distribution is estimated by the average scalewise pooled coefficients, which makes the proposed method computational convenient. The number of features depends on the parametrization of the shearlet system and is equal to the number of scales. For the case of $N = 5$ scales, the methods was compared to similar RR approaches to IQA and was shown to perform better for 5 of 7 distortion categories in the LIVE database and when all categories are considered jointly. RR IQA performance was also compared to a combination of the same multiscale decomposition method with a combination of Weibull distribution fitting and was shown to be superior. This is an interesting result, as the Weibull distribution is not only assumed to model the distribution of edge magnitudes in natural images [24], but its parameter are also shown to be correlated with brain responses measured by EEG [28]. Thus, the Weibull distribution is expected to somewhat better capture properties for modeling images and perception. This could be beneficial as well for EEG-based image quality assessment methods [29]. In order to learn more about statistical image structure and neural processes underlying image perception, future work should identify reasons for the location parameter of the inverse Gaussian distribution to provide a better summary of the multiscale coefficients (at least in the proposed framework for RR IQA). One reason might be that the oriented scale space decompositions evaluated in this work and in [20] are not based on 1st order derivatives and therefore do not comply to the claims in [24]. In this work, the scale parameter λ of the inverse Gaussian distribution has not be considered as a feature, but might further increase the performance of the method.

7. REFERENCES

[1] E. Marieb and K. Hoehn, *Human Anatomy & Physiology*, Pearson Education, 2006.

[2] D. M. Chandler and S. S. Hemami, "VSNR : A Visual Signal-to-Noise Ratio for Natural Images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, 2007.

[3] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[4] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," *IEEE Asilomar Conf. Signals, Syst. Comput.*, vol. 2, no. 1, pp. 1398–1402, 2003.

[5] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A Feature Similarity Index for Image Quality Assess-

ment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, 2011.

[6] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Full-reference image quality assessment using neural networks," in *Int. Work. Qual. Multimed. Exp.*, 2016.

[7] G. Cheng, J. Huang, Z. Liu, and C. Lizhi, "Image quality assessment using natural image statistics in gradient domain," *AEU - Int. J. Electron. Commun.*, vol. 65, no. 5, pp. 392–397, 2011.

[8] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional Neural Networks for No-Reference Image Quality Assessment," in *Comput. Vis. Pattern Recognit. (CVPR), 2014 IEEE Conf.*, 2014, pp. 1733–1740.

[9] M. Saad, A. C. Bovik, and C Charrier, "Blind Image Quality Assessment: A Natural Scene Statistics Approach in the DCT Domain.," *IEEE Trans. image Process.*, vol. 21, no. 8, pp. 3339–3352, 2012.

[10] P. Ye and D. Doermann, "No-Reference Image Quality Assessment Using Visual Codebooks," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3129–3138, 2012.

[11] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *2012 IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 1098–1105.

[12] S. Bosse, D. Maniry, T. Wiegand, and W. Samek, "A deep neural network for image quality assessment," in *Image Processing (ICIP), 2016 IEEE International Conference on*, 2016.

[13] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE image quality assessment database," 2003.

[14] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E. H. Yang, and A. C. Bovik, "Quality-aware images," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1680–1689, 2006.

[15] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: a flexible architecture for multi-scalederivative computation," in *Int. Conf. Image Process.*, 1995, vol. 3, pp. 444–447.

[16] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE J. Sel. Top. Signal Process.*, vol. 3, no. 2, pp. 202–211, 2009.

[17] A. Maalouf, M. C. Larabi, and C. Fernandez-Maloigne, "A grouplet-based reduced reference image quality assessment," in *Int. Work. Qual. Multimed. Exp.*, 2009, pp. 59–63.

[18] M. Carnec, P. Le Callet, and D. Barba, "Objective quality assessment of color images based on a generic perceptual reduced reference," *Signal Process. Image Commun.*, vol. 23, no. 4, pp. 239–256, 2008.

[19] R. Soundararajan and A. C. Bovik, "RRED Indices: Reduced Reference Entropic Differencing for Image Quality Assessment," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 517–526, 2012.

[20] W. Xue and X. Mou, "Reduced Reference Image Quality Assessment based on Weibull Statistics," *Proc. Int. Work. Qual. Multimed. Exp.*, pp. 1–6, jun 2010.

[21] W.-Q Lim, "Nonseparable shearlet transform," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 2056–2065, 2013.

[22] G. Kutyniok and W.-Q Lim, "Compactly supported shearlets are optimally sparse," *J. Approx. Theory*, vol. 163, no. 11, pp. 1564–1589, 2011.

[23] G. Kutyniok and D. Labate, *Shearlets: Multiscale analysis for multivariate data*, Springer Science & Business Media, 2012.

[24] J.-M. Geusebroek, "The Stochastic Structure of Images," *Scale Sp. PDE Methods Comput. Vis.*, pp. 327–338, 2005.

[25] X. Zhang and X. Jing, "Image denoising in contourlet domain based on a normal inverse Gaussian prior," *Digit. Signal Process.*, vol. 20, no. 5, pp. 1439–1446, 2010.

[26] I. Lampl, D. Ferster, T. Poggio, and M. Riesenhuber, "Intracellular Measurements of Spatial Integration and the MAX Operation in Complex Cells of the Cat Primary Visual Cortex," *J. Neurophysiol.*, vol. 92, pp. 2704–2713, 2004.

[27] VQEG, "Objective perceptual assessment of video quality: full reference television," *ITU-T Telecommun. Stand. Bur.*, 2004.

[28] H. S. Scholte and S. Ghebreab, "Brain responses strongly correlate with Weibull image statistics when processing natural images," *J. Vis.*, vol. 9, pp. 1–15, 2009.

[29] S. Bosse, L. Acqualagna, A. K. Porbadnigk, B. Blankertz, G. Curio, K.-R. Müller, and T. Wiegand, "Neurally informed assessment of perceived natural texture image quality," in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1987–1991.