https://doi.org/10.2352/ISSN.2470-1173.2018.12.IQSP-367

No-Reference Image Quality Assessment Using Salient Local Binary Patterns

Pedro Garcia Freitas, Welington Yorihiko Lima Akamine, Mylène Christine Queiroz de Farias; University of Brasília; Brasília, Brazil

Abstract

In this paper, we propose a new no-reference image quality assessment (NR-IQA). The method makes use of local binary patterns (LBP) to label local textures of an image. These labels form a LBP map that can be used to measure the characteristics of image textures (texture map). Then, we compute the histogram of the texture map and weight each LBP label according to its saliency, which is obtained with a visual attention computational model. The weighted histogram is used as input to a regression method that estimates the quality of the image. Experimental results show that the proposed method achieves competitive prediction accuracy and outperforms other state-of-the-art NR-IQA methods. At the same time, the method is simple and reliable, demanding few computational resources, such as memory and processing time.

Introduction

Due to recent advances in multimedia systems, a great effort of the academic community has been dedicated to the design of image quality assessment (IQA) methods that correlate well with the human perception of quality. Quantitative assessment of image quality is an indispensable yet challenging problem in image processing and computer vision research. IQA methods fall into two approaches: subjective and objective. Subjective quality assessment is performed by humans, while objective quality assessment is performed by algorithms designed to mock the subjective judgments. Since quality is thought in terms of human perception, subjective assessment is the utmost reference to describe image quality. However, since it demands humans to perform the scoring tasks, subjective assessments are time-consuming, cumbersome, expensive, and tiresome. In order to overcome these disadvantages, there has been an increasing investment in developing objective methods to measure the image quality in consistent with subjective evaluations.

According to the availability of a “ideal” non-distorted image (reference), objective IQA methods are classified as full reference (FR), reduced-reference (RR) and no-reference (NR). Most accurate these methods determine the quality of images using complex models of the human visual system. Naturally, they are demanding in terms of computational resources [1]. Additionally, IQA methods with the best accuracy performance are generally full-reference IQA (FR-IQA) methods, i.e. methods that require the original content to estimate image quality.

As an alternative, in recent years, several no-reference image quality assessment (NR-IQA) methods have been proposed [2, 3, 4, 5, 6]. Some of the most popular approaches combine image features and machine learning techniques. Some objective image quality assessment methods have been proposed for specific applications and are distortion-specific (DS) methods. Among the state-of-the-art DS-IQA methods, we can cite techniques conceived to assess image sharpness [7, 8], JPEG degradations [9, 10], blockiness artifacts [11, 12], etc. Although these methods can be competent within a scope, they have limited applications in more diversified scenarios.

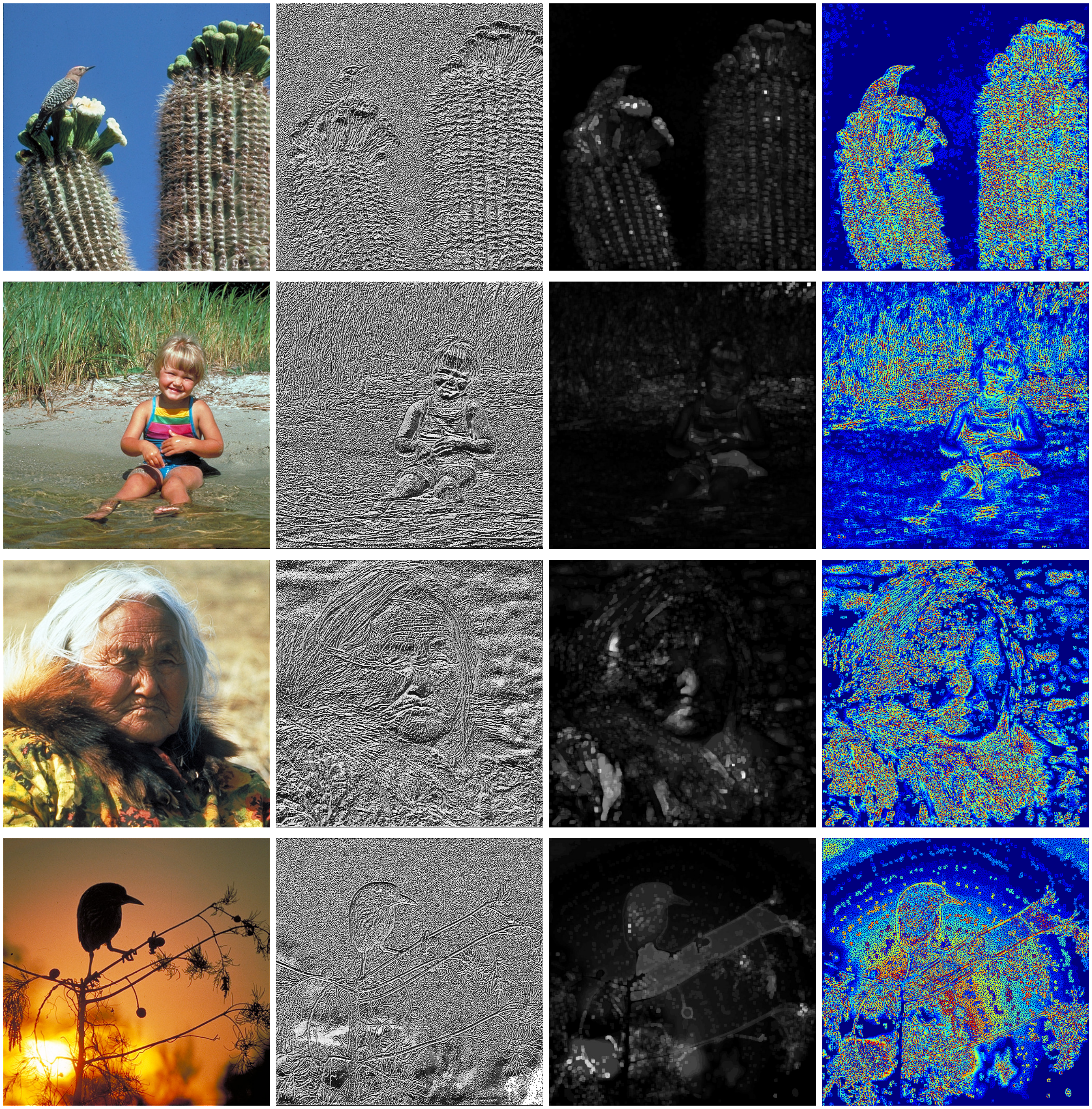
The alternative to DS-IQA methods are general-purpose IQA (GP-IQA) methods, which do not require a prior knowledge about the type of image impairments and are more suitable for diversified scenarios. GP-IQA methods make assumptions about the image characteristics instead of making assumptions about the characteristics of specific impairments. Among the GP-IQA approaches, the methods can be classified as codebook-based [13, 14], ensemble-based [15], or Natural Scene Statistics (NSS)-based approaches [16, 17], as stated by Mittal *et al.* [18]. In codebook-based approaches, the codebook is used to encode basic image features, such as pixel clusters or Gabor filters. In NSS-based approaches, it is assumed that natural images present observable statistic patterns. The quality is measured by computing changes on these statistics of visual naturalnesss. Finally, ensemble-based methods are based on ensemble of regressors, trained on different types of features, such as texture or blur statistics.

Although a better prediction performance is achieved with these state-of-the-art methods, there are still some open questions in the area of NR-IQA [1]. As claimed by Chandler *et al.* [1], NR-IQA methods are focusing on improving prediction accuracy and, so far, few works has been conducted to reduce computational complexity, which is a crucial issue on developing real-time multimedia applications.

In this paper, we propose a GP-NR-IQA method that delivers a good prediction performance and, at the same time, requires limited computational resources. The method uses a machine learning approach that takes into consideration texture features. These texture features are generated using the local binary pattern (LBP) [19] descriptor. This descriptor is chosen based on recent IQA advances which demonstrate that visual impairments affect image textures and their LBP statistics [20, 21, 22, 23]. After generating the texture-based features using the LBP operator, these features are weighted by the saliency of the image regions, which are computed using a visual attention computational model. With a regression learning technique, the weighted features are mapped into a quality score. These combination of texture and saliency generate a new operator, the salient local binary patterns (SLBP), that is proposed in this paper.

Proposed Method

The proposed method is based on two main assumptions. First, visual impairments alter image textures and their statistics. In other words, images with similar impairments, at similar strengths, have textures that share similar statistical properties.



(a) Original ($\mathcal{I}$) (b) LBP ($\mathcal{L}$) (c) BMS ($\mathcal{W}$) (d) SLBP ($\mathcal{S}$)

Figure 1. *Example of original images (a), their LBP (b), BMS (c), and SLBP (d) maps.*

This relation between texture and naturalness measures is used by a few natural scene statistics (NSS) IQA methods [24]. Second, salient visual areas attract more attention than non-salient areas and, therefore, are considered to be more relevant to image quality. Based on this assumption, visual attention models are sometimes used to improve the performance of IQA methods [25].

The texture information of an image $\mathscr{I}$ can be quantified using the uniform local binary pattern (LBP) operator [26], which is defined as:

$$\mathscr{L}[x,y] = LBP^{u}_{R,P}(t_c) = \begin{cases} \sum_{p=0}^{P-1} S(t_p - t_c), & \mathscr{U} \leq 2, \\ P+1, & \text{otherwise}, \end{cases} \quad (1)$$

where

$$\mathscr{U} = |S(t_{P-1} - t_c) - S(t_0 - t_c)| + \sum_{p=1}^{P-1} |S(t_p - t_c) - S(t_{p-1} - t_c)|,$$

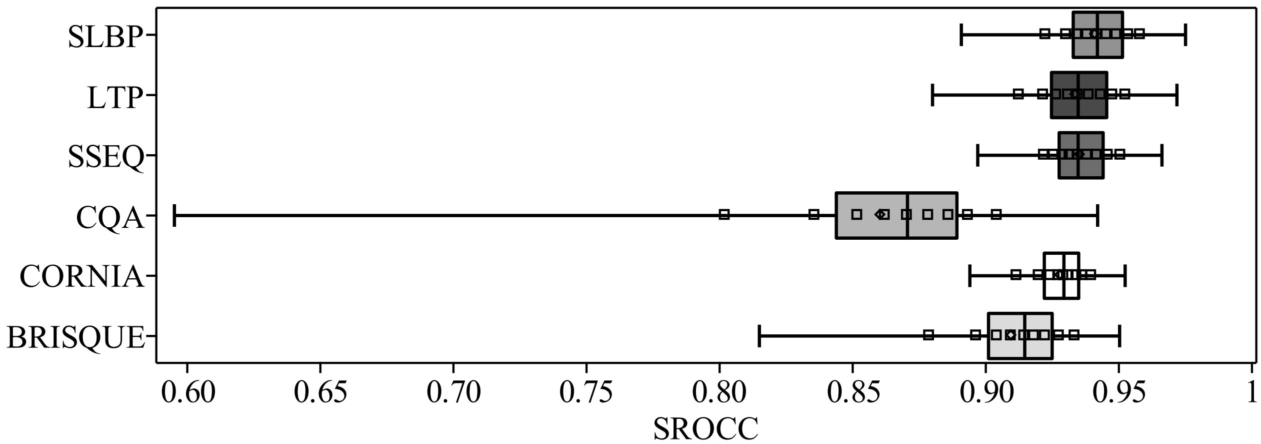


Figure 2. *SROCC distributions from 1,000 random trials on random testing subsets of LIVE2 database.*

P is the total number of neighbors taken into consideration, R is the radius of this neighborhood, and $S(t)$ is a step function given by:

$$S(t) = \begin{cases} 1 & t \geq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

On the above formulation, $t_c = \mathscr{I}[x,y]$ is the value of the pixel of image $\mathscr{I}$ at position (x, y) and $t_p = \mathscr{I}[x_p, y_p]$ is the value of its neighbor, given by:

$$x_p = x + R\cos\left(2\pi\frac{p}{P}\right),$$

and

$$y_p = y - R\sin\left(2\pi\frac{p}{P}\right).$$

To estimate the saliency of the different areas of an image $\mathscr{I}$, we use a computational visual attention model. More specifically, to keep the computational complexity low, we chose the Boolean map-based saliency (BMS) model [27]. When compared with other state-of-the-art visual attention models, BMS is noticeable faster, while still providing a good performance.

After computing the LBP of all pixels of image $\mathscr{I}$, we obtain a LBP map $\mathscr{L}$, where each $\mathscr{L}[x,y]$ gives the local texture associated to the pixel $\mathscr{I}[x,y]$. Similarly, the output of BMS is a saliency map $\mathscr{W}$, where each element $\mathscr{W}[x,y]$ corresponds to the probability that the pixel $\mathscr{I}[x,y]$ attracts the attention of a human observer. The first, second, and third columns of Fig. 1 depict a set of original images $\mathscr{I}$, the corresponding $\mathscr{L}$, and $\mathscr{W}$ maps, respectively.

We generate the feature vector by computing the histogram of $\mathscr{L}$ weighted by $\mathscr{W}$. The histogram $\mathbf{H} = \{h[0], h[1], \cdots, h[P+1]\}$ is given by the following expression:

$$h[\phi] = \sum_i \sum_j \mathscr{W}[i,j] \cdot \Delta(\mathscr{L}[i,j], \phi), \quad (3)$$

where

$$\Delta(v,u) = \begin{cases} 1, & \text{if } v = u, \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

The number of bins of this histogram is similar to the number of distinct LBP patterns of $\mathscr{L}$. So, we can remap each $\mathscr{L}[i,j]$ to its weighted form, generating the map $\mathbb{S}$ displayed in Fig. 1-(d). This figure depicts a heatmap representing the importance of each local texture. We name this weighted LBP map the *salient local binary patterns (SLBP)*.

After computing the feature vector $\mathbf{H}$, we use it to predict the quality of the image $\mathscr{I}$. The prediction is computed using an ensemble-based regression algorithm. We chose the random decision forests (RDF) regression because this technique has been successfully used in pattern recognition applications [28].

Results

We compared the proposed method with the fastest state-of-the-art NR-IQA methods: BRISQUE [4], CORNIA [2], CQA [5], SSEQ [6], and LTP [3]. Given that the performance of NR-IQA methods is measured by how well their predicted scores correlate with the mean observer scores - MOS (scores given by subjects in a psycho-physical experiment), we tested all methods using the LIVE2 database [29]. LIVE2 is a popular image quality database that contains five distortions (JPEG, JPEG2k, Gaussian blue - GB, fast fading - FF, and white noise - WN). We divided the LIVE2 database into two random subsets: 80% for training and 20% for testing. To prevent over-fitting, image contents of the testing subset are not in the training subset, and vice-versa. This 80-20 split is repeated 1,000 times. Table 1 shows the mean SROCC values between predicted scores and MOS values, obtained by averaging 1,000 random trials on random testing subsets. Fig. 2 shows the box plot of the SROCC distribution for each method, over all trials. Additionally, for reference, we also show the results obtained with the FR-IQA methods PSNR and SSIM [30].

In Table 1, the best SROCC results among all methods are shown in italic. The best result among the NR-IQA methods are shown in bold. Notice that, for the complete set of distortions (ALL), the proposed method has the best performance, outperforming all other NR-IQA and FR-IQA methods. When analyzing each type of distortion separately, the proposed method has the best performance among all methods for three of the five distortions. Considering only the NR-IQA methods, the proposed method has the best performance for four of the five distortions.

Table 1. Average SROCC of 1,000 runs of simulations on LIVE2.

Distortion	PSNR	SSIM	BRISQUE	CORNIA	CQA	SSEQ	LTP	SLBP
JPEG	0.8515	*0.9481*	0.8641	0.9002	0.8256	0.9122	0.8961	**0.9211**
JPEG2k	0.8822	0.9438	0.8781	0.9245	0.8366	0.9387	0.9218	***0.9446***
WN	*0.9856*	0.9793	0.9751	0.9501	0.9764	0.9544	0.9363	**0.9808**
GB	0.7818	0.8889	0.9304	0.9465	0.8661	0.9157	0.9256	***0.9598***
FF	0.8869	*0.9335*	0.8468	**0.9132**	0.8488	0.9041	0.8591	0.8708
ALL	0.8013	0.8902	0.9098	0.9281	0.8606	0.9355	0.9195	***0.9414***

To analyze the generalization capability of the tested NR-IQA methods, we performed a cross-database validation test, which consisted of training using the LIVE2 database and testing using the CSIQ [31] and TID2013 [32] databases. Table 2 depicts the SROCC values between predicted scores and MOS values. To perform a straightforward cross-database comparison, only the shared subset of distortions were selected from each database. Notice that the proposed method outperforms the other methods for almost all distortions, with the only exception of the GB distortion in TID2013. Therefore, the cross-database validation test indicates that the proposed method has better generalization properties than the other state-of-the-art IQA methods.

To evaluate the computational cost of the NR-IQA methods, we resized the image shown in Fig 1-(a) from 64x64 to 2048x2048 pixels, creating a set of 8 images with different spatial resolutions (see second column of Table 3). For each image in this set, we computed the feature extraction running time. This process was repeated 100 times and Table 3 shows the average time. Since CQA has restrictions on the minimum spatial resolution required, we could not estimate the extraction running times for resolutions smaller than 256×256. Notice that SLBP has a good time efficiency, loosing only for LTP. But, as depicted in Fig. 3, the proposed method provides a good trade-off between prediction accuracy and computational efficiency.

Conclusions

In this paper, we proposed the salient local binary patterns (SLBP) operator and its use for GP-NR-IQA. Our results show that the proposed SLBP method is a general-purpose NR-IQA, presenting the best prediction performance for the LIVE2 database. SLBP is also generalizable and, therefore, more suitable for general multimedia applications, as can be noticed from the cross-database result tests. The analysis of the computational cost shows that the proposed method presents the best trade-off between accuracy and efficiency, making it adequate for real-time applications. Future works may include a parallel implementation of the SLBP algorithm and an investigation of its suitability for video quality assessment (VQA).

Table 2. SROCC cross-database validation with models trained on LIVE2 and tested on CSIQ and TID2013 databases.

Database	Distortion	BRISQUE	CORNIA	CQA	SSEQ	LTP	SLBP
TID 2013	JPEG	0.8058	0.7423	0.8071	0.7823	0.8472	**0.8946**
	JPEG2k	0.8224	0.8837	0.7724	0.8258	0.9046	**0.9406**
	WN	0.8621	0.7403	0.8692	0.6959	0.6881	**0.8887**
	GB	0.8245	0.8133	0.8214	0.8624	**0.8693**	0.8673
	ALL	0.7965	0.7599	0.8214	0.7955	0.8137	**0.8927**
CSIQ	JPEG	0.8209	0.7062	0.7129	0.8141	0.8784	**0.9130**
	JPEG2k	0.8279	0.8459	0.6957	0.7862	0.8914	**0.9262**
	WN	0.6951	0.8627	0.6596	0.4613	0.7739	**0.9007**
	GB	0.8311	0.8815	0.7648	0.7758	0.8712	**0.9258**
	ALL	0.8022	0.7542	0.7114	0.7403	0.8628	**0.8857**

Table 3. Running time (in seconds) of feature extraction.

Resource	Size	BRISQUE	CORNIA	CQA	SSEQ	LTP	SLBP
Time	16^2	0.0195	1.4622	—	0.0104	0.0003	0.0026
	32^2	0.0185	1.8491	—	0.0159	0.0004	0.0030
	64^2	0.0215	2.8417	—	0.0416	0.0007	0.0085
	128^2	0.0241	5.36	—	0.1279	0.0026	0.0297
	256^2	0.0367	5.2715	—	0.4730	0.0086	0.1378
	512^2	0.0746	5.1854	1.589	1.8259	0.0391	0.4167
	1024^2	0.2440	6.7833	6.1493	7.1955	0.1443	2.1543
	2048^2	1.1599	11.8632	22.7501	29.5196	0.5487	8.1001
Memory	16^2	447.21	3736.28	—	391.41	71.78	113.66
	32^2	447.19	22135.69	—	391.52	72.09	113.25
	64^2	447.21	21163.11	—	436.89	74.19	114.45
	128^2	447.32	33962.16	—	495.88	74.24	124.09
	256^2	781.55	184052.89	—	610.36	76.30	153.21
	512^2	2317.56	1188296.42	1726.22	1098.84	93.18	210.86
	1024^2	8461.55	4921834.67	1725.88	2050.83	149.52	478.71
	2048^2	33041.47	19952894.59	4098.73	32774.72	364.02	1390.57

References

[1] Damon M Chandler. Seven challenges in image quality assessment: past, present, and future research. *ISRN Signal Processing*, 2013, 2013.

[2] Peng Ye, Jayant Kumar, Le Kang, and David Doermann. Unsupervised feature learning framework for no-reference image quality assessment. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1098–1105. IEEE, 2012.

[3] Pedro Garcia Freitas, Welington YL Akamine, and Mylene CQ Farias. No-reference image quality assessment based on statistics of local ternary pattern. In *Quality of Multimedia Experience (QoMEX), 2016 Eighth International Conference on*, pages 1–6. IEEE, 2016.

[4] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012.

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

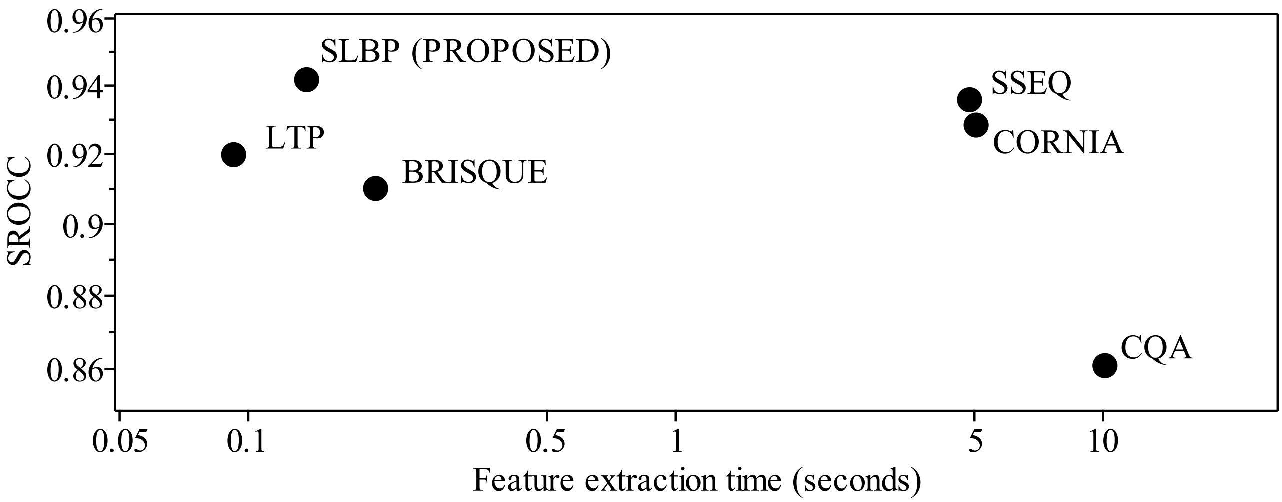


Figure 3. *SROCC versus feature extraction running time.*

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

[7] L. Li, W. Xia, W. Lin, Y. Fang, and S. Wang. No-reference and robust image sharpness evaluation based on multi-scale spatial and spectral features. *IEEE Transactions on Multimedia*, PP(99):1–1, 2016.

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

[9] S Alireza Golestaneh and Damon M Chandler. No-reference quality assessment of jpeg images via a quality relevance map. *Signal Processing Letters, IEEE*, 21(2):155–158, 2014.

[10] Leida Li, Weisi Lin, and Hancheng Zhu. Learning structural regularity for evaluating blocking artifacts in jpeg images. *Signal Processing Letters, IEEE*, 21(8):918–922, 2014.

[11] Leida Li, Yu Zhou, Weisi Lin, Jinjian Wu, Xinfeng Zhang, and Beijing Chen. No-reference quality assessment of deblocked images. *Neurocomputing*, 177:572 – 584, 2016.

[12] Leida Li, Hancheng Zhu, Gaobo Yang, and Jiansheng Qian. Referenceless measure of blocking artifacts by tchebichef kernel analysis. *Signal Processing Letters, IEEE*, 21(1):122–125, 2014.

[13] Peng Ye, Jayant Kumar, Le Kang, and David Doermann. Unsupervised feature learning framework for no-reference image quality assessment. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1098–1105. IEEE, 2012.

[14] Peng Ye and David Doermann. No-reference image quality assessment using visual codebooks. *Image Processing, IEEE Transactions on*, 21(7):3129–3138, 2012.

[15] Huixuan Tang, Neel Joshi, and Ashish Kapoor. Learning a blind measure of perceptual image quality. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 305–312. IEEE, 2011.

[16] Qiaohong Li, Weisi Lin, Jingtao Xu, and Yuming Fang. Blind image quality assessment using statistical structural and luminance features. *IEEE Transactions on Multimedia*, 18(12):2457, 2016.

[17] Lixiong Liu, Yi Hua, Qingjie Zhao, Hua Huang, and Alan Conrad Bovik. Blind image quality assessment by relative gradient statistics and adaboosting neural network. *Signal Processing: Image Communication*, 40:1–15, 2016.

[18] Anish Mittal, Anush K Moorthy, Alan C Bovik, Chang Wen Chen, Periklis Chatzimisios, Tasos Dagiuklas, and Luigi Atzori. No-reference approaches to image and video quality assessment. *Multimedia Quality of Experience (QoE): Current Status and Future Requirements*, page 99, 2015.

[19] Sheryl Brahnam, Lakhmi C Jain, Alessandra Lumini, and Loris Nanni. Introduction to local binary patterns: new variants and applications. In *Local Binary Patterns: New Variants and Applications*, pages 1–13. Springer, 2014.

[20] Min Zhang, Chisako Muramatsu, Xiaoxin Zhou, Tenshi Hara, and Hideaki Fujita. Blind image quality assessment using the joint statistics of generalized local binary pattern. *Signal Processing Letters, IEEE*, 22(2):207–210, 2015.

[21] Jinjian Wu, Weisi Lin, and Guangming Shi. Image quality assessment with degradation on spatial structure. *Signal Processing Letters, IEEE*, 21(4):437–440, 2014.

[22] Yazhong Zhang, Jinjian Wu, Xuemei Xie, and Guangming Shi. Blind image quality assessment based on local quantized pattern. In *Pacific Rim Conference on Multimedia*, pages 241–251. Springer, 2016.

[23] Qiaohong Li, Weisi Lin, and Yuming Fang. Bsd: Blind image quality assessment based on structural degradation. *Neurocomputing*, 236:93–103, 2017.

[24] Anush Krishna Moorthy and Alan Conrad Bovik. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE transactions on Image Processing*, 20(12):3350–3364, 2011.

[25] Mylène CQ Farias and Welington YL Akamine. On performance of image quality metrics enhanced with visual attention computational models. *Electronics letters*, 48(11):631–633, 2012.

[26] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987, 2002.

[27] Jianming Zhang and Stan Sclaroff. Exploiting surroundedness for saliency detection: a boolean map approach. *IEEE transactions on*

pattern analysis and machine intelligence, 38(5):889–902, 2016.

[28] Miao Liu, Mingjun Wang, Jun Wang, and Duo Li. Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and chinese vinegar. *Sensors and Actuators B: Chemical*, 177:970–980, 2013.

[29] HR Sheikh, Z Wang, L Cormack, and AC Bovik. Live image quality assessment database release 2 (2005), 2005.

[30] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[31] Eric C Larson and DM Chandler. Categorical image quality (csiq) database. *Online, http://vision. okstate. edu/csiq*, 2010.

[32] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al. Image database tid2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*, 30:57–77, 2015.

Author Biography

Pedro Garcia Freitas *received a bachelors degree in physics and a masters degree in computer science from the University of Brasília (UnB), Brazil. He is a PhD candidate at the Department of Computer Science of UnB. He also spent one year between 2014 and 2015 as a visiting researcher at Delft University of Technology (TU Delft), in Netherlands. His research interests are in the areas of image and video restoration, programming languages, multimedia quality assessment, and machine learning.*

Welington Yorihiko Lima Akamine *received his BSc degree in computer engineering from the University of Brasília (UnB), Brazil, in 2014 and his M.Sc. degree in electrical engineering from UnB in 2017. His current research interests include visual attention, quality metrics, and image processing.*

Mylène Christine Queiroz de Farias *received her B.Sc. degree in electrical engineering from Universidade Federal de Pernambuco (UFPE), Brazil, in 1995 and her M.Sc. degree in electrical engineering from the Universidade Estadual de Campinas (UNICAMP), Brazil, in 1998. She received her Ph.D. in electrical and computer engineering from the University of California Santa Barbara, USA, in 2004 for work in no-reference video quality metrics. Dr. Farias has worked as a research engineer at CPqD (Brazil) in video quality assessment and validation of video quality metrics. She has also worked for Philips Research Laboratories (The Netherlands) in video quality assessment of sharpness algorithms and for Intel Corporation (Phoenix, USA) developing no-reference video quality metrics. Currently, she is an Assistant Professor in the Department of Electrical Engineering at the University of Brasilia (UnB), where she is a member of the Graduate Program in Informatics (PPGInf) and of the Graduate Program on Electronic Systems and Automation Engineering (PGEA). Dr. Farias is a researcher of the Laboratory of Images, Signals and Audio (LISA) and her current interests include video quality metrics, video processing, multimedia signal processing, watermarking, and information theory. Dr. Farias is a member of IEEE and of the IEEE Signal Processing Society.*