Blind Image Quality Assessment Using the Joint Statistics of Generalized Local Binary Pattern

Min Zhang, *Member, IEEE*, Chisako Muramatsu, Xiangrong Zhou, Takeshi Hara, and Hiroshi Fujita, *Member, IEEE*

***Abstract*—Multimedia, including audio, image, and video, etc., is a ubiquitous part of modern life. Quality evaluation, both objective and subjective, is of fundamental importance for various multimedia applications. In this letter, a novel quality-aware feature is proposed for blind/no-reference (NR) image quality assessment (IQA). The new quality-aware feature is generated from the proposed joint generalized local binary pattern (GLBP) statistics. In this method, using the Laplacian of Gaussian (LOG) filters, the images are first decomposed into multi-scale subband images. Then, the subband images are encoded with the proposed GLBP operator and the quality-aware features are formed from the joint GLBP histograms from the encoding maps of each subband image. Finally, using support vector regression (SVR), the quality-aware features are mapped to the image's subjective quality score for NR-IQA. The experimental results for two representative databases show that the proposed method is strongly correlated to subjective quality evaluations and competitive to the state-of-the-art NR-IQA methods.**

***Index Terms*—Generalized local binary pattern, image quality assessment, no reference, support vector regression.**

I. Introduction

With the increase in the use of multimedia applications in our daily life, perceived quality evaluation has been receiving significant attention. However, accurately modeling and representing a user's reaction to perceived quality remains a challenge. Among these perceived quality evaluation tasks, image quality assessment (IQA) is one of the important issues.

Blind/no-reference (NR) image quality assessment (IQA) has gained considerable importance in the last decade and is considered one of the most difficult quality evaluation tasks. The objective of NR-IQA is to devise perceptual models that can predict the quality of distorted images without any prior information regarding the original image. Most current NR-IQA methods are distortion-specific. They perform NR-IQA quality prediction only if the distortion type of the image is known in advance, for example, blurriness/sharpness or JPEG/JPEG2000 compression distortion etc. [1], [2]. General purpose NR-IQA has been regarded to be a long way to go before reaching useful levels.

Manuscript received March 17, 2014; revised May 13, 2014; accepted May 19, 2014. Date of publication May 22, 2014; date of current version September 10, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Kai-Kuang Ma.

The authors are with the Department of intelligent image information, division of regeneration and advanced medical sciences, Gifu University, Gifu-shi 501-1194, Japan (e-mail: dr.mzhang1980@gmail.com)

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2014.2326399

Most of the present general-purpose NR-IQA algorithms employ natural scene statistics (NSS), which is based on the hypothesis that natural scenes possess certain statistical properties that are altered in the presence of distortion, rendering them unnatural. It is considered that by characterizing this un-naturalness of NSS; one can identify the distortion affecting the image and perform NR-IQA. Statistical models to characterize the NSS-based NR-IQA algorithms have been investigated in both the spatial domain such as BRISQUE [3] and the transform domain such as wavelet [4], [5] and DCT domain [6]. They evaluate the "unnaturalness" extent of the test image without any prior knowledge.

Another category of NR-IQA algorithms is based on the learned regression model for the description of a set of relevant features affecting the quality [7]–[9]. Features used in the training stage could be the selected raw image patches in the given images (e.g., LBIQ in [7]), phase congruency features and the entropy of the distorted image (e.g., GRNN in [8]), and image patches learned from a visual codebook by K-means (e.g., CBIQ in [9]). However, the quality-aware feature extraction and selection are crucial in these approaches.

Previously, the image local structure has proved to be reasonably simple and useful in the shaping of full reference (FR) IQA approaches [10], [11]. From the bottom-up point of view of the human vision system (HVS), image-structure primitives in the early visual cortex, area V1, play a crucial role in the perception of quality [12]. When an image deteriorates, the local structures and their statistical distribution vary. Shaping an IQA strategy from local structure primitives eliminates most of the information redundancy in the images. Therefore, this type of IQA index has a significant potential in the application of reduced reference (RR) and NR-IQA.

The NR-IQA method proposed in this letter attempts to mimic the statistical model of most representative local structure features in the early vision stage that affect quality. Preliminary related work has been undertaken with a type of low-level local structure descriptor, the local binary pattern (LBP) [13], in our recent work [14], [15]. LBP can be considered as the binary approximation of the image structure primitives in the early stage of HVS. Hence, the statistics of local structure primitives in the early vision stage can be formed easily. The LBP operator labels the pixels of an image by thresholding the n-neighborhood of each pixel with the center pixel value. Each local structure pattern is encoded as a binary string from the sign component of thresholding result. However, the performance of LBP is bounded to some extent as a local structure descriptor since it lacks of magnitude information. Hence, there is still considerable room for improvement.

This letter mainly includes two contributions. First, a new local structure descriptor called the generalized local binary

pattern (GLBP) is proposed. Second, the joint statistics of the GLBP are employed as the quality-aware feature to form a novel blind image quality assessment algorithm. The experimental results on two large-scale image databases indicate that the proposed NR-IQA algorithm is effective and performs statistically better than other state-of-the-art NR-IQA approaches, as well as PSNR and the FR measure SSIM [10].

II. Blind Image Quality Assessment using the Joint Statistics of the Generalized Local Binary Pattern

The proposed quality-aware feature extraction for blind image quality assessment has the following procedures.

A. Multi-scale decomposition

From a computational point of view, the response of the classical cortical receptive field (CRF) in the HVS can be modeled using a series of hierarchical filters [12], i.e., the Laplacian of Gaussian (LOG) filters, which is defined as

$$\nabla^2 G(r,\sigma) = -\frac{1}{\pi\sigma^4}\left(1-\frac{r^2}{2\sigma^2}\right)\exp\left(-\frac{r^2}{2\sigma^2}\right), \quad (1)$$

where $G(r,\sigma)=\frac{1}{2\pi\sigma^2}\exp(-\frac{r^2}{2\sigma^2})$ is the Gaussian kernel, r is the l_2 norm of $\boldsymbol{X}$,i.e., $r=\|\boldsymbol{X}\|_2$, and $\boldsymbol{X}$ is the position of the test image $\boldsymbol{I}$. The filter response $\boldsymbol{F}_{Subband}$ of image $\boldsymbol{I}$ can be obtained by the convolution of the filter kernel $\nabla^2 G(r,\sigma)$ and the image $\boldsymbol{I}$ as

$$\boldsymbol{F}_{Subband}(\boldsymbol{X},\sigma) = \nabla^2 G(r,\sigma)\otimes \mathbf{I}(\boldsymbol{X}). \quad (2)$$

B. Generalized LBP and Image Perceptual Quality

The traditional LBP operator [13] takes the form

$$LBP_{P,R}(t_c)=\sum_{p=0}^{P-1} S(t_p-t_c)2^p, \text{ where } S(t)=\begin{cases}1, & t\geq 0\\ 0, & t<0\end{cases} \quad (3)$$

In this letter, the proposed generalized LBP is defined as

$$GLBP_{P,R,T}(t_c)=\sum_{p=0}^{P-1} S'(t_p-t_c)2^P,$$
$$\text{where} \quad S'(t)=\begin{cases}1, & t\geq T\\ 0, & t<T.\end{cases} \quad (4)$$

P denotes the total number of involved neighbors and R denotes the radius of the neighborhood. At the center pixel t_c in the image, the generalized LBP code is defined in the following form: each neighboring pixel t_p is assigned with a binary label of either "0" or "1," depending on whether the difference between the center pixel t_c and the neighboring pixel t_p is greater than the set threshold T. GLBP becomes LBP when $T=0$.

Similar to the definition of the uniform LBP, the "uniform" pattern of the GLBP is defined as a uniformity measure U ("pattern").

$$U(GLBP_{P,R,T}) = |S'(t_{P-1}-t_c)| + \sum_{p=1}^{P-1}|S'(t_p-t_c)-S'(t_{p-1}-t_c)| \quad (5)$$

A U value of two or less is defined as a "uniform" pattern. It refers to the patterns that have limited transitions or discontinuities (bitwise 0/1 changes) in the GLBP code.

A local rotation invariant pattern of the GLBP could be defined as

$$GLBP^{riu2}_{P,R,T}=\begin{cases}\sum_{p=0}^{P-1} S'(t_p-t_c) & \text{if} U(GLBP_{P,R})\leq 2\\ P+1 & \text{otherwise}\end{cases} \quad (6)$$

The superscript ^{riu2} reflects the use of the rotation invariant "uniform" patterns that have a U value of two or less.

With the proposed GLBP operator, the GLBP encoding map for each subband image is defined as

$$GLBP_{P,R,\sigma,T}(\boldsymbol{X}_c)=\sum_{p=0}^{P-1} S'(\boldsymbol{F}(\boldsymbol{X}_p,\sigma)-\boldsymbol{F}(\boldsymbol{X}_c,\sigma))2^P. \quad (7)$$

Then, the normalized histogram from the GLBP encoding map of each subband image is represented as follows:

$$\boldsymbol{H}_{GLBP_{P,R,\sigma,T}}(k)=\sum_{\mathbf{X}} f(GLBP_{P,R,T,\sigma}(\boldsymbol{F}(\boldsymbol{X},\sigma)),k)/(MN),$$
$$\text{where} \quad f(x,y)=\begin{cases}1, x=y\\ 0, \text{otherwise}\end{cases} \quad \text{and} \quad k\in[0,K]. \quad (8)$$

where K is the maximum GLBP pattern value and the subband image $\boldsymbol{F}_{Subband}(\boldsymbol{X},\sigma)$ is assumed to be of size of $M\times N$.

For each subband image, we stack the histograms $\boldsymbol{H}_{GLBP_{P,R,\sigma,T}}$ with multi-threshold T to represent the subband image

$$\boldsymbol{J}^W_\sigma=(\boldsymbol{H}_{GLBP_{P,R,\sigma,T_0}},\boldsymbol{H}_{GLBP_{P,R,\sigma,T_1}},\ldots,\boldsymbol{H}_{GLBP_{P,R,\sigma,T_{W-1}}}), \quad (9)$$

where W is the total number of the selected parameter T.

Finally, the quality-aware feature $\boldsymbol{J}$of the entire image $\boldsymbol{I}$ is formed from the joint features $\boldsymbol{J}^W_\sigma$ from each subband.

$$\boldsymbol{J}_{GLBP}(\mathbf{I})=(\boldsymbol{J}^W_{\sigma_0},\boldsymbol{J}^W_{\sigma_1},\ldots,\boldsymbol{J}^W_{\sigma_{D-1}}). \quad (10)$$

D corresponds to the number of image decomposition scales.

Fig. 1 shows an example of the GLBP histogram comparison of two distortion types with different image content in a single decomposition scale. As can be seen from Fig. 1, it is both very interesting and apparent that images with the same distortion type and close distortion levels share considerably similar GLBP histogram profiles despite their different content. This is the principal motivation of our method.

However, statistics of the GLBP from a single parameter T may suffer from instability and lack of the robustness to be the discriminative feature for quality evaluation. The joint GLBP from introduced multi-threshold provides extra information from the description of sensibility of the GLBP code against different parameter T. This additional information can contribute to the discrimination of different local structures, especially in middle to high frequency subbands.

C. Perceptual Quality Prediction

In the quality prediction stage, a nonlinear regression model, support vector regression (SVR), has been widely adopted in NR-IQA [4], [9], [15]. SVR is a learning process and is highly effective in high-dimensional data pooling [16]. With SVR, the

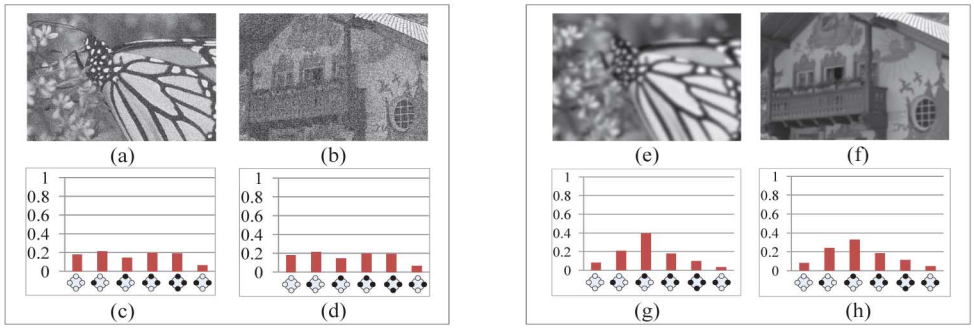

Fig. 1. Comparison of uniform and rotation invariant GLBP histograms calculated from the same decomposed scale with $R = 1$, $P = 4$ (a) Gaussian noise contained image from "monarch.bmp" in the LIVE database (cropped for visibility), (b) Gaussian noise contained image from "paintedhouse.bmp" in the LIVE database (cropped for visibility), (c) Uniform and rotation invariant GLBP histogram of the image in (a), (d) Uniform and rotation invariant GLBP histogram of the image in (b), (e) Gaussian blurred image from "monarch.bmp" in the LIVE database (cropped for visibility), (f) Gaussian blurred image from "paintedhouse.bmp" in the LIVE database (cropped for visibility), (g) Uniform and rotation invariant GLBP histogram of the image in (e), (h) Uniform and rotation invariant GLBP histogram of the image in (f).

TABLE I
MEDIAN SROCC COMPARISON ACROSS 1000 TRAIN-TEST COMBINATIONS ON THE LIVE DATABASE (THE TWO BEST CORRELATIONS ARE MARKED IN ITALIC)

	Type	Ref. No. for *Train.*	Feature Domain	JP2K	JPEG	WN	Blur	FF	All
PSNR	FR	\	Spatial	0.8646	0.8831	0.9410	0.7515	0.8736	0.8636
SSIM			Spatial	0.9389	0.9466	0.9635	0.9046	0.9393	0.9129
CBIQ	NR	23	Gabor	0.8935	0.9418	0.9582	0.9324	0.8727	0.8954
LBIQ			Wavelet	0.9040	0.9291	0.9702	0.8983	0.8222	0.9063
BLIINDS-II (SVM)			DCT	*0.9285*	0.9422	0.9691	0.9231	*0.8893*	0.9306
DIIVINE			Wavelet	0.913	0.910	*0.984*	0.921	0.863	0.916
LD-TS			Wavelet	0.8202	0.8334	0.9566	0.9251	0.8863	0.8833
BRISQUE			Spatial	0.9139	*0.9647*	*0.9786*	*0.9511*	0.8768	*0.9395*
NR-GLBP			LOG	*0.9465*	*0.9563*	*0.9786*	*0.9537*	*0.8890*	*0.9511*

joint GLBP histograms are mapped to the quality score to train the quality prediction model.

$$Q_{predict} = SVR(\boldsymbol{J}_{GLBP}, \mathcal{MOD}) \tag{11}$$

where $\mathcal{MOD}$ is the trained model for regression and $Q_{predict}$ is the objective quality score predicted by the model.

III. EXPERIMENTAL SETUP

The LIVE IQA database [17] and TID2008 database [18] are used to test the performance of the proposed method. The LIVE database consists of 29 reference images with 779 distorted images, including five different distortion categories. The TID2008 database includes 25 reference images distorted using 17 types of distortions. We only test on 4 common types with the LIVE database. Besides, the 25th reference image and the related distorted images are excluded because it's not natural but synthesized geometry texture.

Since the proposed approach requires a training procedure, the LIVE database is split into two randomly chosen subsets: 80% training and 20% testing. There is no overlap between the training and testing data sets in each split and evaluation. The LIBSVM package [19] is utilized to implement the SVR. In our experimentations, we perform ε-SVR with a radial-basis function (RBF) kernel. The parameters (cost, gamma) of ε-SVR are tuned by a two-dimensional (2D) grid search in the logarithm space. The random train-test procedure is repeated 1000 times and the median of the performance across these 1000 iterations is reported to eliminate the performance bias. The proposed NR method is compared with other NSS-based NR-IQA methods including DIIVINE [4], LD-TS [5], BLIINDS [6], LBIQ [7], visual codebook-based metric (CBIQ) [9], and spatial domain metric BRISQUE [3].

In this letter, the proposed method is indicated as NR-GLBP. The training and testing images are decomposed into four scales. The scales σ of these filters are selected as [0.5 1.3 2.6 5.2]. $R = 1$, $P = 4$ are selected for computational simplicity and data rate compaction. Three sets of fixed parameter $\boldsymbol{T} = [-1, 0, 6]$ are used. For each subband image with each parameter T, six-dimensional quality-aware features from the uniform and rotation invariant GLBP operator are generated. This result in 18 scalar features for each subband image. The total number of quality-aware features of an image is 72.

The Spearman's rank ordered correlation coefficient (SROCC) and Pearson's (linear) correlation coefficient (LCC) between the predicted score from the algorithm and human opinion are used to assess the IQA performance. Before computing the LCC, the algorithm scores were passed through a logistic nonlinearity as described in [11].

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A preliminary evaluation of the proposed method is undertaken in terms of correlation with human perception. Tables I and II summarize both the performance of the overall prediction accuracy and the specific distortion prediction accuracy of the proposed algorithm with other state-of-the-art algorithms, as well as two FR indexes - PSNR and SSIM [10] on the LIVE database. Table III summarizes the comparison of the performance of the proposed algorithm with that of the competitive NR-IQA methods on the TID2008 database.

From Tables I and II, it is evident that statistics from the low-level structure features of images are highly correlated to human perceptual quality. The proposed algorithm is competitive with other state-of-the-art NR-IQA methods on LIVE database. Further, it is statistically superior to the full-reference PSNR and

TABLE II
MEDIAN LCC COMPARISON ACROSS 1000 TRAIN-TEST COMBINATIONS ON THE LIVE DATABASE (THE TWO BEST CORRELATIONS ARE MARKED IN ITALIC)

	Type	Ref. No. for *Train.*	Feature Domain	JP2K	JPEG	WN	Blur	FF	All
PSNR	FR	\	Spatial	0.8762	0.9029	0.9173	0.7801	0.8795	0.8592
SSIM			Spatial	0.9405	0.9462	0.9824	0.9004	0.9514	0.9066
CBIQ	NR	23	Gabor	0.8898	0.9454	0.9533	0.9338	0.8951	0.8955
LBIQ			Wavelet	0.9103	0.9345	0.9761	0.9104	0.8382	0.9087
BLIINDS-II (SVM)			DCT	*0.9348*	0.9676	0.9799	0.9381	0.8955	0.9302
DIIVINE			Wavelet	0.922	0.921	*0.988*	0.923	0.888	0.917
LD-TS			Wavelet	0.8507	0.9544	0.9315	0.8784	0.8774	0.8273
BRISQUE			Spatial	0.9229	*0.9734*	0.9851	*0.9506*	*0.9030*	*0.9424*
NR-GLBP			LOG	*0.9559*	*0.9716*	*0.9853*	*0.9542*	*0.9122*	*0.9542*

TABLE III
MEDIAN SROCC PERFORMANCE COMPARISON ACROSS 1000 TRAIN-TEST COMBINATIONS ON THE TID2008 DATABASE (THE TWO BEST CORRELATIONS ARE MARKED IN ITALIC)

	Type	JP2K	JPEG	WN	Blur	All
PSNR	FR	0.8762	0.9029	0.9173	0.7801	0.8592
SSIM		0.9405	0.9462	0.9824	0.9004	0.9066
DIIVINE	NR	*0.924*	0.866	*0.851*	0.862	0.889
BLIINDS-II		0.9157	0.8901	0.6600	0.8500	0.8442
BRISQUE		0.832	*0.924*	0.829	*0.881*	*0.896*
NR-GLBP		*0.9504*	*0.9325*	*0.8889*	*0.9023*	*0.9399*

SSIM. From Table III, the proposed algorithm performs statistically better than other NR approaches to IQA in both the individual distortion types and overall performance on the TID2008 database.

From Marr's theory [12], the local structural primitives in the early vision stage are crucial to represent image semantic information in late HVS process. We consider that the information existed in the basic primitive signals can be approximated by the proposed GLBP operator. Since the proposed algorithm attempts to extract the statistics of local structure primitives that the natural scenes share in common as the quality-aware feature instead of any specific distortion features (such as blocking), this facilitates the proposed algorithm to a broad range of distortion measurement and stable across different databases. However, these are only the preliminary results under current implementations; further research is needed to clarify the underlying mechanism and improve the performance.

Due to the simplicity of the GLBP operator, the data rate of the proposed algorithm is low and the average computational time for quality-aware feature extraction on the LIVE database is only 1.151 second/image using a laptop with an Intel i5 processor at 2.60 GHz with un-optimized MATLAB code.

V. CONCLUSION

This letter proposed a new local structure descriptor called GLBP to form a general-purpose blind IQA index. The tentative experimental results on two subject-rated image databases showed that the proposed NR-IQA index exhibits good correlation with subjective evaluation over a wide variety of image distortions with a low data rate and high efficiency. It outperforms the popular FR methods, including PSNR and SSIM index, and is competitive to the state-of-the-art NR-IQA methods.

ACKNOWLEDGMENT

The authors gratefully thank the members of Fujita Lab. for the valuable discussions. Further, the authors acknowledge the Venture Business Laboratory (VBL) of Gifu University for their support in this research work. The authors are also grateful to the editor, the anonymous reviewers, and Dr. Yi Fang for their valuable comments.

REFERENCES

[1] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 717–728, 2009.
[2] H. R. Sheikh, A. C. Bovik, and L. K. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1918–1927, 2005.
[3] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
[4] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.
[5] X. Gao, X. Li, D. Tao, L. He, and W. Lu, "Universal no reference image quality assessment metrics based on local dependency," in *IEEE ACPR*, 2011.
[6] 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.
[7] H. Tong, M. Li, H. Zhang, C. Zhang, J. He, and W. Ma, "Learning no-reference quality metric by examples," in *Proc. 11th Int. Multimedia Modelling Conf.*, Jan. 2005, pp. 247–254.
[8] C. Li, A. C. Bovik, and X. Wu, "Blind image quality assessment using a general regression neural network," *IEEE Trans. Neural Netw.*, vol. 22, pp. 793–799, 2011.
[9] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebook," in *IEEE International Conference on Image Processing*, 2011.
[10] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error measurement to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.
[11] M. Zhang, X. Mou, and L. Zhang, "Non-Shift edge based Ratio (NSER): An Image Quality Assessment Metric based on Early Vision Features," *IEEE Signal Process. Lett.*, vol. 18, no. 5, pp. 315–318, May 2011.
[12] D. Marr, *Vision*. San Francisco, CA, USA: W.H. Freeman, 1980.
[13] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002.
[14] M. Zhang, X. Mou, H. Fujita, L. Zhang, X. Zhou, and W. Xue, "Local binary pattern statistics feature for reduced reference image quality assessment," *IS&T/SPIE Electronic Imaging (Int. Soc. Optics and Photonics)*, pp. 86600L)–86600L-8, 2013.
[15] M. Zhang, J. Xie, X. Zhou, and H. Fujita, "No reference image quality assessment based on local binary pattern statistics," *Vis. Commun. Image Process. (VCIP)*, pp. 1–6, 2013.
[16] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, 2004.
[17] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, *LIVE Image Quality Assessment Database Release 2* [Online]. Available: http://live.ece.utexas.edu/research/quality
[18] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008-a database for evaluation of full-reference visual quality assessment metrics," *Adv. Mod. Radioelectron.*, vol. 10, no. 10, pp. 30–45, 2009.
[19] C. Chang and C. Lin, *LIBSVM: A library for support vector machines*, 2011 [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm/