Blind Image Quality Assessment Using Singular Value Decomposition Based Dominant Eigenvectors For Feature Selection

Besma Sadou[1], Atidel Lahoulou[2*], Toufik Bouden[1], Anderson R. Avila[3], Tiago H. Falk[3], Zahid Akhtar[4]

[1]Non Destructive Testing Laboratory, University of Jijel, Algeria
[2]LAOTI laboratory, University of Jijel, Algeria
[3]Institut National de la Recherche Scientifique, University of Québec, Montreal, Canada [4]University of Memphis, USA

Abstract

In this paper, a new no-reference image quality assessment (NR-IQA) metric for grey images is proposed using LIVE II image database. The features used are extracted from three well-known NR-IQA objective metrics based on natural scene statistical attributes from three different domains. These metrics may contain redundant, noisy or less informative features which affect the quality score prediction. In order to overcome this drawback, the first step of our work consists in selecting the most relevant image quality features by using Singular Value Decomposition (SVD) based dominant eigenvectors. The second step is performed by employing Relevance Vector Machine (RVM) to learn the mapping between the previously selected features and human opinion scores. Simulations demonstrate that the proposed metric performs very well in terms of correlation and monotonicity.

Keywords

Natural Scene Statistics (NSS), Singular Value Decomposition (SVD), dominant eigenvectors, Relevance Vector Machine (RVM).

1. Introduction

In the present decade, no reference image quality assessment (NR-IQA) and enhancement has become an interesting topic in image processing as it handles the image without the need for its original version which may not exist in some applications (e.g., image restoration). Indeed, the most efficient NR-IQA metrics are based on Natural Scene Statistics (NSS) which assume that all original images are natural and that the distortions disrupt this naturalness and make images seem unnatural [1]. This fact may make users feel uncomfortable with visual data and may consequently affect their judgement concerning data's visual quality. Most of the commonly used NR-IQA metrics in the literature are based on NSS features which are extracted from different domains such as the Discrete Wavelet Transform (DWT) domain (e.g. BIQI [2] and DIIVINE [3]), the Discrete Cosine Transform (DCT) domain (e.g. BLIINDS [4] and BLIINDS-II [5]) and

Natarajan Meghanathan et al. (Eds) : CCSEA, CLOUD, SIPRO, AIFU, SEA, DKMP, NCOM - 2019
pp. 233-242, 2019. © CS & IT-CSCP 2019 DOI: 10.5121/csit.2019.90919

the spatial domain (e.g. BRISQUE [6] and NIQE [7]). However, the latest generation of metrics exploit the multi-domain information which simulates well the hierarchical structure of the visual cortex perception [8-9] (e.g. WG-LAB [10] and metrics proposed in [11] and [12]).

In the present paper, the first step of the framework is to extract a descriptive vector containing natural scene statistics features from multiple domains, namely DWT, DCT and spatial domain. For the quality estimator to be accurate, the descriptive vector should hold as less as possible of generated features which can be relevant, irrelevant or redundant. Since the work by A. Lahoulou et al. [13], more and more researchers apply feature selection methods to image quality assessment in order to keep only informative features that describe better the visual quality attributes.

Feature selection models can be classified into three main categories [14] : (1) Filter models: where a relevance index is calculated for each feature independently of the predictor considering some measures such as information measure, (2) Wrappers: these methods use learning algorithms to identify relevant features. This is what makes it more accurate than filter methods but time consuming and computationally expensive, and (3) Embedded models: it is a combination of the two previous methods, the feature selection is embedded in the learning process e.g. decision trees.

In this paper, we develop a new and efficient NR-IQA metric for grey level images. First, a features vector is extracted using three well known NR-IQA metrics operating in three different domains (i.e. DCT domain, DWT domain and spatial domain) in order to better capture human vision properties.

After that, the variable selection process is launched to keep only the most pertinent attributes. This step is performed by using an embedded method namely the dominant eigenvectors after the singular value decomposition (SVD). Finally, the nonlinear regression algorithm of the relevance vector machine (RVM) is applied to generalize prediction of quality scores to out of sample images. The LIVE (release 2) image quality database [15] provides the ground truth data (i.e. the DMOS values) as well as the test images from which the features vector is computed.

2. FEATURES EXTRACTION AND SELECTION

2.1. Image Features Extraction

The features used in this paper come from three learning-based NR-IQA metrics namely BRISQUE, BIQI and BLIINDS-II summarized in table 1 below. The size of the vectors of features is 36, 18 and 24, respectively. The blind metrics where these features come from are described as follows:

2.1.1. BRISQUE [6]

This metric does not require any transformation of the image. It directly extracts NSS features in the spatial domain. For each image, a generalized Gaussian distribution (GGD) is used to estimate the distribution, and then generates the parameters as resulted features. 18 features are extracted using 2 scales, resulting in 36 features used to evaluate the perceptual quality of an image.

2.1.2. BIQI [2]

This algorithm is based on the extraction of NSS in the wavelet domain over three scales and three orientations. Three features are extracted (mean, variance and shape) and used to classify a distorted image into one of N distortions using support vector machine (SVM), then support vector regression (SVR) is used to predict quality score.

2.1.3. BLIINDS-II [5]

Presented by Saad et al., this model works in the DCT domain. A total of 24 features are extracted from the block DCT domain and are affected by changing the type and the level of distortion. These features are then input to the Bayesian inference model to get the perceived quality estimate.

Table 1. NR-IQA metrics considered to investigate the relevance of features for perceptual quality judgement..

NR-IQA algorithm	Domain	Features	
BRISQUE	Spatial domain	36	$f_1, \ldots, f_{36}$
BIQI	DWT domain	18	$f_{37}, \ldots, f_{54}$
BLIINDS-II	DCT domain	24	$f_{55}, \ldots f_{78}$

As a first step, we build in a 78-D vector of original attributes by putting all the extracted features together.

2.2. Feature Selection Technique

All 78 descriptors previously discussed are extracted from LIVE image database release 2 (LIVE II) [15]. This database contains 29 high resolution colour reference images degraded by 5 distortion types (JPEG2000, JPEG, white noise, Gaussian blur, and transmission errors using a fast fading Rayleigh channel model). A set of 982 test images is subjectively evaluated by 29 observers and the Difference Mean Opinion Scores (DMOS) are calculated as recommended by the Video Quality Experts Group (VQEG) Phase I FR-TV [16]. DMOS corresponds to the difference of the Mean Opinion Scores between reference and distorted images.

In order to eliminate redundant and irrelevant features and select only useful ones, we used singular value decomposition (SVD), which is one among a large array of techniques used for dimension reduction.

SVD decomposes a M(m x n) matrix into three matrices as:

$$M = USV^T \text{ (Eq. 1)}$$

where U and V are two orthogonal matrices of (m x p) and (n x p) dimensions, respectively.
S is a (p x p) diagonal matrix.
p is called the rank of matrix M
The diagonal positive entries of matrix S are called singular values of M. These values are arranged in descending order of their magnitude.

For feature selection, we used the same algorithm as the column select problem [17]. This algorithm can be summarized in the following steps:

i. Input the matrix where rows are images and columns are features.
ii. Compute the centralized data.
iii. Apply SVD to get the main components.
iv. Get the dimensions having most of the variation (select the dominant eigenvectors, e.g. representing the 95% of the data).
v. Compute leverage scores using the dominant eigenvectors of the principal components (.i.e. the norm of the eigenvector's coefficients).
vi. Sort the leverage scores in descending order.
vii. Get the indices of the vectors with the largest leverage scores.

In this paper, we select the features which have a leverage score greater than or equal to 0.4. Figure 1 shows the resulting selected features with their leverage scores. We can note that the most significant features come from BLIINDS-II no-reference quality metric.

3. PREDICTION MODEL

In this work, Relevance Vector Machine (RVM) [18] is employed as prediction model instead of support vector machine (SVM) [19, 20] which is the most common. This choice is made based on the benefits the RVMs offer over the SVMs, mainly probabilistic predictions and automatic estimation of the hyper-parameters.

For a given set of samples $\{x_i, t_i\}_{i=1}^{N}$ where x_i is the input variable vector, t_i is the target value, N is the length of training data. The RVM regression expression is:

$$t(x) = \sum_{i=1}^{N} w_i K(x, x_i) + w_0 + \varepsilon_n \qquad \text{(Eq. 2)}$$

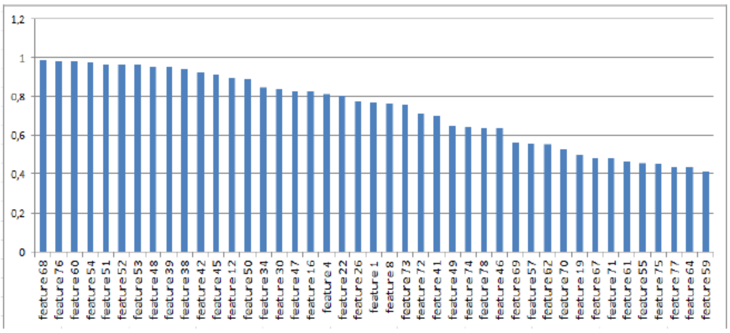

Figure 1. The most significant features with their leverage scores

Where N is the number of data points, $w = [w_1,.., w_n]$ is weights vector, w_0 is the bias, $K(x,x_i)$ is the kernel function and $\varepsilon_n = N(0,\sigma^2)$ is the error term with zero Gaussian mean and variance σ^2. Usually, the Gaussian Kernel is preferred and its formula is:

$$K(x, x_i) = exp\left[\frac{-(x-x_i)^T(x-x_i)}{2S^2}\right] \quad \text{(Eq. 3)}$$

Where S^2 is the Gaussian kernel width.

Assuming that the samples $\{x_i, t_i\}_{i=1}{}^N$ are independently generated, the likelihood of all data set can be written as follows:

$$P(t|w, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} exp\left\{-\frac{1}{2\sigma^2}\|t - \varphi_w\|^2\right\} \quad \text{(Eq. 4)}$$

Where φ is a design matrix having the size $N * (N+1)$ with:

$$\varphi(x_i) = [1, K(x_i, x_1), K(x_i, x_2), .., K(x_i, x_N)]^T \quad \text{(Eq. 5)}$$

The highest probability estimation of w and σ^2 of equation (4) may suffer from serious over-fitting. To solve this, Tipping [18] imposed an explicit zero-mean Gaussian prior probability distribution for the weights, w, with diagonal covariance of α as follows:

$$P(w|\alpha) = \prod_{i=0}^{N} N\left(w_i \middle| 0, \alpha_i^{-1}\right) \quad \text{(Eq. 6)}$$

Where α is a vector of $(N + 1)$ named hyper parameters.

In this way, using Baye's rule, the posterior over all unknown parameters could be calculated given the defined non informative prior distribution:

$$P(w, \alpha, \sigma^2|t) = \frac{P\left(t\middle|w, \alpha, \sigma^2\right).P(w,\alpha,\sigma)}{\int P\left(t\middle|w, \alpha, \sigma^2\right)P(w,\alpha,\sigma^2)\, dw\, d\alpha\, d\sigma^2} \quad \text{(Eq. 7)}$$

Full analytical solution of the integral of (Eq. 7) is obdurate. Thus, decomposition of the posterior distribution according to equation 8 below is called upon to ease the solution [18].

$$P(w, \alpha, \sigma^2|t) = P(w|t, \alpha, \sigma^2)P(\alpha, \sigma^2|t) \quad \text{(Eq. 8)}$$

The posterior distribution over the weights is calculated using Bayes rule and is given by:

$$P(w|t, \alpha, \sigma^2) = \frac{P\left(t\middle|w, \sigma^2\right)P(w|\alpha)}{P\left(t\middle|\alpha, \sigma^2\right)} \quad \text{(Eq. 9)}$$

The resulting posterior distribution over the weights is the multivariate Gaussian distribution:

$$P(w|t, \alpha, \sigma^2) = N(\mu, \epsilon) \quad \text{(Eq. 10)}$$

Where the mean and the covariance are respectively expressed by:

$$\mu = \sigma^{-2}\epsilon\varphi^T t \quad \text{(Eq. 11)}$$

$$\epsilon = (\sigma^{-2}\varphi^T\varphi + A)^{-1} \quad \text{(Eq. 12)}$$

With diagonal matrix $A = diag(\alpha_0, .., \alpha_N)$

For uniform hyper priors over α and σ^2, one requires only to maximize the term $(t|\alpha, \sigma^2)$ as follows:

$$P(t|\alpha,\sigma^2) = \int P(t|w,\sigma^2) P(w,\alpha)\, dw \quad \text{(Eq. 13)}$$

$$P(t|\alpha,\sigma^2) = \left[(2\pi)\frac{-N}{2} / \sqrt{|\sigma^2 + \varphi A^{-1}\varphi^T|}\right] * exp\left\{-\frac{1}{2} t^T (\sigma^2 + \varphi A^{-1}\varphi^T)^{-1} t\right\} \quad \text{(Eq. 14)}$$

By simply forcing the derivatives of Equation (14) to zero, we can get the re-estimation formulas on for α and σ^2 respectively as follow:

$$\alpha_i^{new} = \frac{1-\alpha_i \epsilon_i}{\mu_i^2} \quad \text{(Eq. 15)}$$

$$(\sigma^2)^{new} = \frac{\|t - \varphi_\mu\|^2}{N - \sum_i (1 - \alpha_i \epsilon_i)} \quad \text{(Eq. 16)}$$

4. EXPERIMENTS AND RESULTS

The performance of our metric is evaluated using two criteria: Pearson Correlation Coefficient (PCC) and Spearman Rank Order Correlation Coefficient (SROCC) between subjective and objective scores. The first criterion gives estimation about the prediction linear correlation while the second measures the prediction monotonicity.

Before computing PCC, a nonlinear mapping between true DMOS and algorithm scores is carried out using the logistic function with five parameters [21]. The expression of the quality score which is the predicted MOS is given by:

$$DMOS_p = \beta_1 logistic(\beta_2, D - \beta_3) + \beta_4 D + \beta_5 \quad \text{(Eq. 17)}$$

Where D and $DMOS_p$ are the predicted scores before and after regression, respectively.

β_1 to β_5 are the regression parameters estimated using fmin search function in Matlab's optimization Toolbox. The logistic function is given by:

$$logistic(\tau, D) = \frac{1}{2} - \frac{1}{1+exp(\tau D)} \quad \text{(Eq. 18)}$$

We randomly split the images of LIVE II database into two non-overlapping sets, 80% for training and the remaining 20% for test phase. This random splitting is repeated 100 times in order to ensure the robustness of our metric. At the end, we calculate the average of the obtained performance criteria.

Tables 2 and 3 give SROCC and PCC mean values between subjective and objective scores on each of the five distortion subsets and the entire database (noted by ALL). These values are compared to six state-of-the-art general-purpose NR-IQA metrics (BIQI, BLIINDS, DIIVINE, BLIINDS-II, BRISQUE and NIQE).

Table 2. SROCC of different methods on LIVE II database.

	JP2K	JPEG	WN	Gblur	FF	ALL
BIQI	0.736	0.591	0.958	0.778	0.700	0.726
BLIINDS	0.805	0.552	0.890	0.834	0.678	0.663
DIIVINE	0.913	0.910	**0.984**	0.921	0.863	0.916
BLIINDS -II	0.951	0.942	0.978	0.944	**0.927**	0.920
BRISQUE	0.914	**0.965**	0.979	**0.951**	0.877	0.940
NIQE	0.917	0.938	0.966	0.934	0.859	0.914
Proposed	**0.949**	0.924	0.982	0.946	0.884	**0.955**

Table 3. PCC of different methods on LIVE II database.

	JP2K	JPEG	WN	Blur	FF	ALL
BIQI	0.750	0.630	0.968	0.800	0.722	0.740
BLIINDS	0.807	0.597	0.914	0.870	0.743	0.680
DIIVINE	0.922	0.921	**0.988**	0.923	0.888	0.917
BLIINDS-II	**0.963**	**0.979**	0.985	0.948	**0.944**	0.923
BRISQUE	0.923	0.974	0.985	0.951	0.903	0.942
NIQE	0.937	0.956	0.977	0.953	0.913	0.915
Proposed	0.962	0.945	0.981	**0.957**	0.911	**0.953**

Numerical results show that the proposed no-reference image quality assessment metric achieves good performances in terms of monotonicity (table 2) and correlation (table 3). The first position best results are mentioned in bold whereas the second positions best results are the underlined values. We can notice that the proposed metric gives the first or the second best performance for all the subsets except that of the encoded images via JPEG algorithm.

Furthermore, the scatter plot of our method for test set with median SROCC is provided in figure 2; where the horizontal axis corresponds to the objective scores and the vertical axis corresponds to subjective scores. Every dot in the plot represents an image in the database. It can be seen that most of the dots are clustered around the red line that represent ideal linear correlation line "Proposed=DMOS", this means that the proposed metric achieves good correlation with human scores.

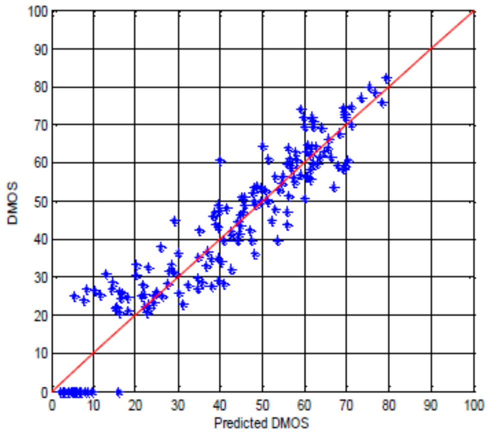

Figure 2. The scatter plots of the predicted perceived quality vs. the DMOS

5. CONCLUSION

The two main ideas of this work is that the most successful NR-IQA metrics are based on NSS features and that the multi-domain information simulate well the hierarchical structure of the visual cortex perception. For these reasons, the features used to build the present NR-IQA metric are collected from three NR-IQA methods based on NSS features operating in three different domains (spatial, DWT and DCT). Only pertinent features are input to the relevance vector machine algorithm to predict the objectives score. The step of feature selection is achieved using Singular Value Decomposition (SVD) based dominant eigenvectors. Numerical experiments show that the proposed metric is competitive with DIVINE, BLIINDS II and BRISQUE methods. It also outperforms BLIINDS, BIQI and NIQE algorithms.

REFERENCES

[1] D. Zhang, Y. Ding , N. Zheng, "Nature scene statistics approach based on ICA for no- reference image quality assessment", Proceedings of International Workshop on Information and Electronics Engineering (IWIEE), 29 (2012), 3589- 3593.

[2] A. K. Moorthy, A. C. Bovik, A two-step framework for constructing blind image quality indices[J], IEEE Signal Process. Lett., 17 (2010), 513-516.

[3] L. Zhang, L. Zhang, A.C. Bovik, A Feature-Enriched Completely Blind Image Quality Evaluator, IEEE Transactions on Image Processing, 24(8) (2015), 2579- 2591.

[4] M.A. Saad, A.C. Bovik, C. Charrier, A DCT statistics-based blind image quality index, Signal Process. Lett. 17 (2010) 583–586.

[5] M. A. Saad, A. C. Bovik, C. Charrier, Blind image quality assessment: A natural scene statistics approach in the DCT domain, IEEE Trans. Image Process., 21 (2012), 3339-3352.

[6] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (2012), 4695 - 4708.

[7] A. Mittal, R. Soundararajan, A. C. Bovik, Making a completely blind image quality analyzer, IEEE Signal Process. Lett., 20 (2013), 209-212.

[8] N. Kruger, P. Janssen, S. Kalkan, M. Lappe, A. Leonardis, J. Piater, A. Rodriguez-Sanchez, L. Wiskott, "Deep hierarchies in the primate visual cortex: What can we learn for computer vision?", IEEE Trans. Pattern Anal. Mach. Intell., 35 (2013), 1847–1871.

[9] D. J. Felleman, D. C. Van Essen, “Distributed hierarchical processing in the primate cerebral cortex,” Cerebral cortex, 1 (1991), 1–47.

[10] B. Sadou, A. Lahoulou, T. Bouden, A New No-reference Color Image Quality Assessment Metric in Wavelet and Gradient Domains, 6th International Conference on Control Engineering and Information Technologies, Istanbul, Turkey, 25-27 October (2018), 954-959.

[11] Q. Wu, H. Li, F. Meng, K. N. Ngan, S. Zhu, No reference image quality assessment metric via multi-domain structural information and piecewise regression. J. Vis. Commun. Image R., 32(2015), 205–216.

[12] X. Shang, X. Zhao, Y. Ding, Image quality assessment based on joint quality-aware representation construction in multiple domains, Journal of Engineering 2018 (2018), 12p.

[13] A. Lahoulou, E. Viennet, A. Beghdadi, ‘‘Selecting low-level features for image quality assessment by statistical methods,’’ J. Comput. Inf. Technol. CIT 18 (2010), 83–195.

[14] H. Liu, H. Motoda, R. Setiono, and Z. Zhao, “Feature Selection: An Ever Evolving Frontier in Data Mining”, Journal of Machine Learning Research, Proceedings Track, pp. 4-13, 2010.

[15] H. R. Sheikh, Z. Wang, L. Cormack, A. C. Bovik, LIVE Image Quality Assessment Database Release 2, http://live.ece.utexas.edu/research/quality

[16] Final VQEG report on the validation of objective quality metrics for video quality assessment:http://www.its.bldrdoc.gov/vqeg/projects/frtv_phaseI/

[17] M. W. Mahoney, P. Drineas, “CUR matrix decompositions for improved data analysis,” in Proc. the National Academy of Sciences, February 2009.

[18] M.E. Tipping. The relevance vector machines. In Advances in Neural Information Processing Systems 12, Solla SA, Leen TK, Muller K-R (eds). MIT Press: Cambridge, MA (2000), 652-658.

[19] D. Basak, S. Pal, D.C. Patranabis, Support vector regression, Neural Information Processing – Letters and Reviews, 11 (2007).

[20] B. SchÖlkopf, A.J. Smola, Learning with Kernels. MIT press, Cambridge, (2002).

[21] H. R. Sheikh, M. F. Sabir, A. C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, IEEE Trans. Image Process., 15 (2006), 3440–3451.

AUTHORS

Besma Sadou is currently a PhD student in the department of Electronics at university of Jijel (Algeria). She also works as full-time teacher of mathematics at the middle school. Her research interests are focused on reduced and no-reference image quality assessment.

Atidel Lahoulou is Doctor in Signals and Images from Sorbonne Paris Cité (France) since 2012. She earned her Habilitation Universitaire in 2017 and is currently associate professor in the department of computer science at university of Jijel (Algeria). Her research interests include visual data quality evaluation and enhancement, biometrics, machine learning and cybersecurity.

Toufik Bouden received the engineer diploma (1992), MSc (1995) and PhD (2007) degrees in automatics and signal processing from Electronics Institute of Annaba University (Algeria). Since 2015, he is full professor in the department of Automatics. His areas of research are signal and image processing, non-destructive testing and materials characterization, biometrics, transmission security and watermarking, chaos, fractional system analysis, synthesis and control.

Anderson R. Avila received his B.Sc. in Computer Science from Federal University of Sao Carlos, Brazil, in 2004 and his M.Sc in Information Engineering from Federal University of ABC in 2014. In October 2013, Anderson worked as a short-term visiting researcher at INRS, where he now pursues his Ph.D degree on the topic of speaker and emotion recognition. His research interests include pattern recognition and multimodal signal processing applied to biometrics.

Tiago H. Falk is an Associate Professor at INRS-EMT, University of Quebec and Director of the Multimedia Signal Analysis and Enhancement (MuSAE) Lab. His research interests are in multimedia quality measurement and enhancement, with a particular focus on human-inspired technologies.

Zahid Akhtar is a research assistant professor at the University of Memphis (USA). Prior to joining the University of Memphis, he was a postdoctoral fellow at INRS-EMT-University of Quebec (Canada), University of Udine (Italy), Bahcesehir University (Turkey), and University of Cagliari (Italy), respectively. Dr. Akhtar received a PhD in electronic and computer engineering from the University of Cagliari (Italy). His research interests are biometrics, affect recognition, multimedia quality assessment, and cybersecurity.