No Reference Image Quality Metric via Distortion Identification and Multi-Channel Label Transfer

Qingbo Wu*, Hongliang Li*, King N. Ngan†*, Bing Zeng*‡ and Moncef Gabbouj§
*Institute of Image Processing, University of Electronic Science and Technology of China, Chengdu, China
†Department of Electronic Engineering, The Chinese University of Hong Kong, ShaTin, Hong Kong
‡Department of Electronic and Computer Engineering,
The Hong Kong University of Science and Technology, Clearwater Bay, Kowloon, Hong Kong
§Department of Signal Processing, Tampere University of Technology, Tampere, Finland
Email: wqb.uestc@gmail.com; hlli@uestc.edu.cn; knngan@ee.cuhk.edu.hk; eezeng@uestc.edu.cn; moncef.gabbouj@tut.fi

***Abstract*—In this paper, we propose a no reference image quality assessment (NR-IQA) algorithm based on distortion identification (DI) and multi-channel label transfer (LT). First, the distortion type classification is used to obtain the query image's probabilities of belonging to each distortion type. Then, the distortion specific label transfer is implemented in multiple distortion category channels. Based on the hypothesis that the similar images share the similar subjective qualities, the label transfer predicts the subjective quality of the query image by pooling the labels of its *k*-nearest neighbors (KNN) retrieved from the annotated samples. A weighting average of the multi-channel label transfer's outputs is computed to obtain the final perceptual quality score. The weight is the query image's probability that belongs to the corresponding distortion type. The experimental results show that the proposed method outperforms representative NR-IQA approaches and some full-reference metrics.**

I. Introduction

No reference image quality assessment (NR-IQA) is an important and challenging task in the image processing. The existing NR-IQA approaches are mainly composed of two steps: quality-aware feature extraction and predictive model regression/learning.

Many spatial domain, transform domain and saliency information [1], [2] were exploited in extracting the quality-aware features. As widely discussed in [3]–[5], the distribution of the natural images' coefficients in the transform domain presents obvious regularity. In these methods, the natural scene statistics (NSS) of an image is usually represented by a parametric statistical model. Meanwhile, the NSS features are only extracted from one channel of the color image. There are two issues in this feature extraction procedure. First, the approximating from the NSS to a parametric model will induce fitting error, which may reduce the discriminant power of the NSS. Second, the single-channel feature can't capture a comprehensive perceptual artifact for the color image.

In addition, to learn the mapping from the quality-aware features to the subjective quality score, the support vector regression or neural network methods are employed in many NR-IQA methods. To guarantee the robustness, these metrics usually require lots of training samples. Meanwhile, the black-box mapping makes the relationship between the these features and the perceptual quality implicit.

To address the problems mentioned above, we propose a novel NR-IQA method based on distortion identification and label transfer (DILT). Inspired by the nonparametric scene parsing method proposed in [6], we consider the IQA as an image annotation problem, where the ground-truth label is the subjective quality score (e.g., DMOS or MOS). Based on the hypothesis that the similar images share the similar subjective qualities, the k-nearest neighbors (KNN) of the query image are retrieved from some annotated samples. Then, the labels of the KNN are used for predicting the perceptual quality with a distance based pooling scheme. Here, we directly compute the statistical information among each frequency band to represent the frequency domain features. All of these features are extracted from the YCbCr channels for a color image.

To further improve the label transfer accuracy, we introduce the distortion identification module into the proposed method. Then, multiple distortion specific label transfer channels are built, where the specific feature combinations are assigned to each channel. By computing the weighting average of all channels' outputs, we can obtain the final predicted quality score. The weight is the probability of the query image that belongs to the corresponding distortion type. For clarity, the framework of our proposed method is shown in Fig. 1, where $p_1 \sim p_N$ denote the query image's probabilities of belonging to each distortion type, $Q_1 \sim Q_N$ denote the predicted quality score in each label transfer channel and Q denote the final predicted quality metric.

The remainder of this paper is organized as follows. Section II describes the proposed quality-aware features and Section III presents the proposed quality evaluation metric. The experimental results are shown in Section IV. Finally, we summarize our paper in Section V.

II. Quality-aware features extraction

A. DCT domain

In this paper, we compute the DCT coefficients for all of the non-overlapped $W \times W$ blocks in an image, where the W is set to 8. Each line along the down-left $45°$ direction is defined as the frequency band f in the coefficients block. Let $y_{j,f}(i)$ denote the ith DCT coefficient of the jth block in the frequency band f. The normalized frequency band coefficient $c_{j,f}$ in the fth band and jth block can be given by

$$c_{j,f} = \sqrt{\frac{1}{N_f}\sum_{i=1}^{N_f} y_{j,f}^2(i)} \quad (1)$$

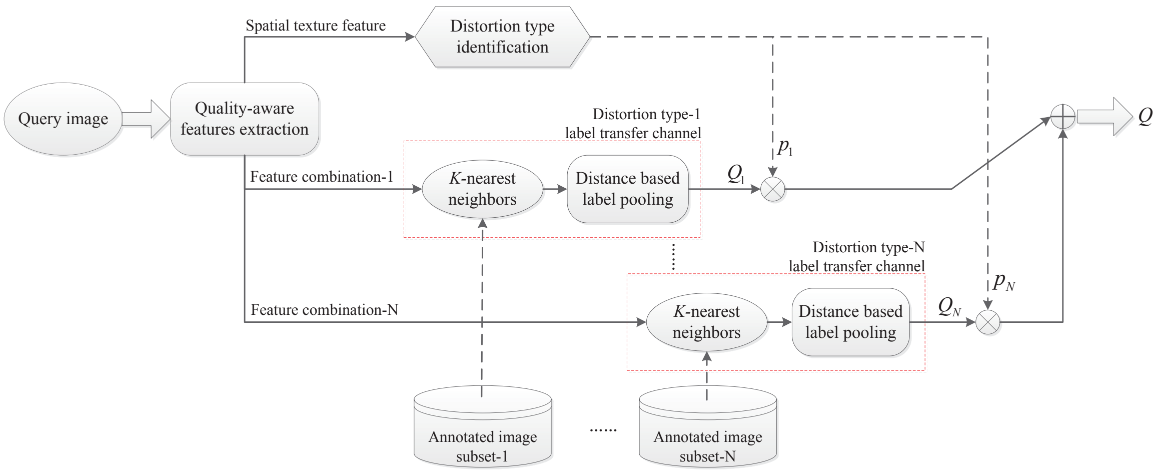

Fig. 1. The framework of our proposed DILT method.

where N_f is the number of DCT coefficients in the fth frequency band.

To quantitatively measure the energy compaction property in the DCT domain, three features are extracted from the DCT coefficients, i.e., frequency band entropy, inter-band difference entropy and the distribution of skewness.

Firstly, the fth band's entropy e_f in the DCT domain is defined as

$$e_f = -\sum_{i=1}^{N_e} p_f^e(i) \log_2 p_f^e(i) \tag{2}$$

where $p_f^e(i)$ is the probability of the ith bin in the margin distribution associated with the fth DCT frequency band across all blocks of $c_{j,f}$. N_e is the number of the bins in partitioning the DCT frequency band coefficients.

Correspondingly, the frequency band entropy feature can be obtained by combining the band entropy in all of frequency bands except for the DC coefficient, i.e.,

$$E_f = [e_1, e_2, \cdots, e_{N_f}] \tag{3}$$

where N_f is the number of high frequency bands, which is 14 for the 8×8 block. In our proposed feature extraction scheme, the same operation is implemented in all of the YCbCr channels. Then, we can obtain the multi-channel feature,

$$E_{MCF} = [E_f^Y, E_f^{Cb}, E_f^{Cr}] \tag{4}$$

Secondly, we further compute the entropy for the difference of the neighboring frequency bands' coefficients. Let $g_{j,f}$ denote the difference of the neighboring frequency bands' coefficients, i.e.,

$$g_{j,f} = c_{j,f} - c_{j,f+1} \tag{5}$$

The entropy of $g_{j,f}$ across all blocks can be defined as

$$d_f = -\sum_{i=1}^{N_d} p_f^d(i) \log_2 p_f^d(i) \tag{6}$$

where $p_f^d(i)$ is the probability of the ith bin in the margin distribution associated with $g_{j,f}$ across all blocks. N_d is the number of the bins. By combining the inter-band difference entropy in all neighboring frequency bands, the second DCT domain feature can be obtained, i.e.,

$$D_f = [d_1, d_2, \cdots, d_{N_f - 1}] \tag{7}$$

Similarly, we define the YCbCr channel feature D_{MCF} as

$$D_{MCF} = [D_f^Y, D_f^{Cb}, D_f^{Cr}] \tag{8}$$

Thirdly, the distribution of the skewness across all blocks is computed to measure the asymmetry of the coefficients. Let c_j denote the frequency band coefficient variable in the $j$the block ($c_{j,f} \in c_j$) and s_j denote its skewness, which is given by

$$s_j = \frac{E(c_j - \mu(c_j))^3}{\sigma^3(c_j)} \tag{9}$$

where $\mu(c_j)$ and σ are the mean value and standard deviation of c_j, respectively. Let s denote the margin distribution of the skewness across all blocks. Then, the YCbCr channel feature S_{MCF} can be given by,

$$S_{MCF} = [s^Y, s^{Cb}, s^{Cr}] \tag{10}$$

B. Wavelet domain

In the wavelet domain, we exploit its exponential decay and self-similarity properties across all scales. Firstly, we employ the entropy of each subband to quantitatively measure the exponential decay property. Here, each channel of a color image is decomposed into L scales with the wavelet transform, where L is set to 4. Let $\hat{e}_{k,l}$ denote the wavelet coefficient entropy of the kth direction under the lth scale, where $1 \le l \le L$ and $k = \{1, 2, 3\}$ correspond to the HL, LH and HH directions, respectively. Then, the definition of $\hat{e}_{k,l}$ can be given by

$$\hat{e}_{k,l} = -\sum_{i=1}^{N_w} \hat{p}_{k,l}(i) \log_2 \hat{p}_{k,l}(i) \tag{11}$$

where $\hat{p}_{k,l}(i)$ is the probability of the ith bin in the margin distribution associated with the kth direction under the lth scale. N_w is the number of bins in partitioning the wavelet coefficients.

Since the exponential decay is along with the scales variation, we represent each channel's wavelet entropy feature by cascading all subband entropies in three directions, i.e.,

$$\begin{aligned}\hat{e}_{HL} &= [\hat{e}_{1,1}, \hat{e}_{1,2}, \cdots, \hat{e}_{1,L}]\\ \hat{e}_{LH} &= [\hat{e}_{2,1}, \hat{e}_{2,2}, \cdots, \hat{e}_{2,L}]\\ \hat{e}_{HH} &= [\hat{e}_{3,1}, \hat{e}_{3,2}, \cdots, \hat{e}_{3,L}]\\ \hat{e} &= [\hat{e}_{HL}, \hat{e}_{LH}, \hat{e}_{HH}]\end{aligned} \tag{12}$$

By combining the entropy features in YCbCr channels, the multi-channel fusion wavelet entropy feature E_{MCF} can be given by

$$\hat{E}_{MCF} = [\hat{e}^{Y}, \hat{e}^{Cb}, \hat{e}^{Cr}] \tag{13}$$

Secondly, we use the Kullback-Leibler divergence (KLD) to quantitatively measure self-similarity property in the wavelet domain. Let $\hat{d}_{k,l}$ denote the neighboring subbands' KLD between the lth and the $(l+1)$th scales along the kth direction, where $1 \le l \le L-1$ and k share the same meaning in (11). Then, the definition of KLD can be given by

$$\hat{d}_{k,l} = \sum_{i=1}^{N_w} \hat{p}_{k,l}(i) \log_2 \Big(\frac{\hat{p}_{k,l}(i)}{\hat{p}_{k,l+1}(i)} \Big) \tag{14}$$

where a larger $\hat{d}_{k,l}$ indicates weaker similarity between two neighboring subbands.

The KLD feature is also represented by cascading the $\hat{d}_{k,l}$ in three directions, i.e.,

$$\begin{aligned}\hat{d}_{HL} &= [\hat{d}_{1,1}, \hat{d}_{1,2}, \cdots, \hat{d}_{1,L-1}]\\ \hat{d}_{LH} &= [\hat{d}_{2,1}, \hat{d}_{2,2}, \cdots, \hat{d}_{2,L-1}]\\ \hat{d}_{HH} &= [\hat{d}_{3,1}, \hat{d}_{3,2}, \cdots, \hat{d}_{3,L-1}]\\ \hat{d} &= [\hat{d}_{HL}, \hat{d}_{LH}, \hat{d}_{HH}]\end{aligned} \tag{15}$$

Similarly, the multi-channel fusion KLD feature $\hat{D}_{MCF}$ can be defined as,

$$\hat{D}_{MCF} = [\hat{d}^{Y}, \hat{d}^{Cb}, \hat{d}^{Cr}] \tag{16}$$

C. Spatial domain

Except for the frequency features discussed above, we also extract the texture feature in the spatial domain. Here, we employ the classic local binary pattern (LBP) [7] descriptor to extract the spatial feature. Unlike the features discussed above, the binary descriptor only capture the gradient directions' distribution, which is lack of the gradient intensity information. In extracting the feature from the Cb or Cr channel, many perceptual insensitive structural differences will be preserved in the LBP feature, which may reduce the discriminability of the LBP feature. Thus, we only extract the LBP feature from the single channel gray image.

III. PREDICTION MODEL

In this section, we describe the proposed two-step prediction model, which consists of distortion type identification and label transfer.

Firstly, we divide the annotated images into different subsets according to their distortion types. Each of the distortion specific subsets is used to construct a label transfer channel. Here, we identify the distortion type of the query image with the off-line trained SVM classifier. In our experiment, the input of the classifier is the single-channel LBP feature and the outputs are the query image's probabilities of belonging to each distortion type.

Secondly, the distortion specific label transfer is separately implemented in each label transfer channel for the query image. In our distance based label transfer criterion, the KNN of the query image is retrieved from the distortion specific image subset. By computing the weighted average of the KNN's labels, we can obtain the predicted quality score in current label transfer channel.

Let F_i^q denote the ith quality-aware feature vector extracted from the query image and $F_{i,j}^r$ denote the ith quality-aware feature vector of a reference image in the jth annotated image subset. Here, we define the overall feature distance H of two images as the product of the chi-square distance for each pair of the feature vectors, i.e.,

$$\begin{aligned}H_j &= \prod_{i=1}^{N_F} h(F_i^q, F_{i,j}^r)\\ h(F_i^q, F_{i,j}^r) &= \sum_{t=1}^{N_i} \frac{(F_i^q(t) - F_{i,j}^r(t))^2}{F_i^q(t) + F_{i,j}^r(t)}\end{aligned} \tag{17}$$

where $F_i^q(t)$ and $F_{i,j}^r(t)$ denote the tth element of the feature vectors F_i^q and $F_{i,j}^r$, N_F is the number of the features selected for current distortion type and N_i is the dimension of the ith feature.

In our experiment, we search for the top five nearest neighbors of the query image from current annotated image subset in terms of the feature distance H_j. When we received the transferred labels (i.e., DMOS) from the k-nearest neighbors, a distance based weighting scheme is used to predict the subjective quality in current label transfer channel. The normalized weight $w_j(u)$ for the uth selected neighbor image under jth distortion type can be defined as

$$w_j(u) = H_j^{-1}(u) / \sum_{u=1}^{5} H_j^{-1}(u) \tag{18}$$

Then, the predicted quality index in the jth label transfer channel can be computed as

$$Q_j = \sum_{u=1}^{5} w_j(u) \cdot DMOS_j(u) \tag{19}$$

By assigning larger weight to the quality index which has more similar distortion type with the query image, we can obtain the final predicted perceptual quality as

$$Q = \sum_{j=1}^{N} p_j \cdot Q_j \tag{20}$$

where p_j is the query image's probability of belonging to the jth distortion type and N is the number of the distortion types in the reference image set.

TABLE I. Feature combinations for different distortion specific label transfer channels

LT channel ID	Distortion type	Feature combination
1	JPEG2000	$S_{MCF}, \hat{E}_{MCF}, \hat{D}_{MCF}$, LBP
2	JPEG	$D_{MCF}, S_{MCF}, \hat{E}_{MCF}, \hat{D}_{MCF}$, LBP
3	WN	$E_{MCF}, \hat{E}_{MCF}, \hat{D}_{MCF}$, LBP
4	Blur	$D_{MCF}, \hat{E}_{MCF}, \hat{D}_{MCF}$, LBP
5	FF	$S_{MCF}, \hat{E}_{MCF}, \hat{D}_{MCF}$, LBP

According to the empirical analysis, we elaborately select some specific features which produce the most accurate KNN for each distortion specific label transfer channel. The detailed feature combinations are shown in Table I.

IV. Experimental Results

To verify the performance of our proposed DILT algorithm, we conduct the experiments on the widely used LIVE II IQA databases [8]. In the consistency experiment, we randomly split all of the images in LIVE II into two non-overlapped sets, i.e., the training set and the testing set. The training set is composed of 80% of the reference images and the distorted versions of them. The testing set consists of the rest 20% of reference images and their distorted images.

In our proposed algorithm, the training set serves two purposes, i.e., learning distortion types classifier and constructing the annotated image set for the label transferring. In order to ensure the robustness of our proposed metric, we conduct the random splitting evaluation 100 times on the LIVE II database. The median of the indexes across the 100 trails is used for the performance of our proposed method. The Spearman's rank ordered correlation coefficient (SROCC) between the predicted quality metric Q and the true DMOS is adopted to measure the performance of different NR-IQA approaches.

Some representative full reference image quality assessment (FR-IQA) (e.g., PSNR, SSIM [9]) and NR-IQA (e.g., BLINDS [3], DIIVINE [5] and BLINDS-II [4]) methods are involved in the comparison. The detailed consistency performance comparison is presented in Table II. It can be seen that although the proposed DILT method can't perform best under each single distortion type, we still achieve suboptimal performance. For the general purpose NR-IQA algorithm, the predicted perceptual quality should work well for all of the distortion types. In this principle, our proposed DILT method outperforms all of the remaining NR-IQA algorithms in the entire database, whose SROCC can be up to 0.934. Furthermore, the proposed DILT method is even superior to the FR-IQA algorithms PSNR (SROCC=0.820) and SSIM (SROCC=0.851). Furthermore, the intuitive scatter plot result of the proposed method is shown in Fig. 2. It can be seen that the predicted quality index Q is obviously monotonic and consistent with the DMOS value for each plot.

V. Conclusion

In this paper, we propose a novel two-step NR-IQA algorithm based on distortion identification and label transfer. Unlike existing single-channel feature and support vector regression combination methods, we extract the quality-aware features from YCbCr channels and elaborately design a nonparametric distortion specific label transfer for predicting the perceptual quality. In the comparison experiment, our proposed DILT method shows obvious superiority over some representative NR-IQA algorithms and even outperforms some classic FR-IQA approaches.

TABLE II. Median SROCC across 100 train-test trials

Distortion		JPEG2000	JPEG	WN	Blur	FF	All
Metric	Type						
PSNR	FR	0.890	0.841	0.985	0.782	0.890	0.820
SSIM	FR	0.932	0.903	0.963	0.894	0.941	0.851
BLINDS	NR	0.805	0.552	0.890	0.834	0.678	0.663
DIIVINE	NR	0.913	0.910	0.984	0.921	0.863	0.916
BLINDS-II	NR	0.951	0.942	0.978	0.944	0.927	0.920
DILT	NR	0.898	0.923	0.979	0.940	0.903	0.934

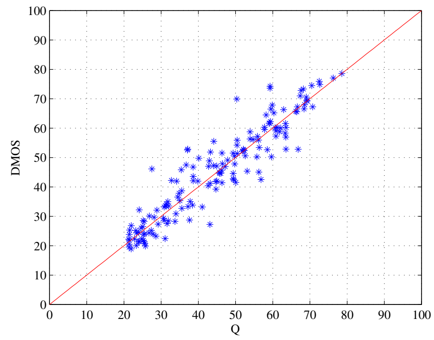

Fig. 2. The scatter plots of the predicted quality index Q vs. the DMOS.

Acknowledgments

This work was partially supported by NSFC (No. 61271289), National High Technology Research and Development Program of China (863 Program, No. 2012AA011503), The Ph.D. Programs Foundation of Ministry of Education of China (No. 20110185110002) and The Fundamental Research Funds for the Central Universities (ZYGX2012YB007).

References

[1] H. Li, L. Xu, and G. Liu, "Two-layer average-to-peak ratio based saliency detection," *Signal Processing: Image Communication*, vol. 28, no. 1, pp. 55 – 68, 2013.

[2] H. Li and K. N. Ngan, "A co-saliency model of image pairs," *Image Processing, IEEE Transactions on*, vol. 20, no. 12, pp. 3365–3375, 2011.

[3] M. Saad, A. Bovik, and C. Charrier, "A dct statistics-based blind image quality index," *IEEE Signal Processing Letters*, vol. 17, no. 6, pp. 583–586, 2010.

[4] M. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the dct domain," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

[5] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[6] C. Liu, J. Yuen, and A. Torralba, "Nonparametric scene parsing via label transfer," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2368–2382, 2011.

[7] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[8] 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.

[9] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.