Blind Image Quality Assessment Based on Multichannel Feature Fusion and Label Transfer

Qingbo Wu, *Member, IEEE*, Hongliang Li, *Senior Member, IEEE*, Fanman Meng, *Member, IEEE*, King N. Ngan, *Fellow, IEEE*, Bing Luo, Chao Huang, and Bing Zeng, *Senior Member, IEEE*

***Abstract*— In this paper, we propose an efficient blind image quality assessment (BIQA) algorithm, which is characterized by a new feature fusion scheme and a *k*-nearest-neighbor (KNN)-based quality prediction model. Our goal is to predict the perceptual quality of an image without any prior information of its reference image and distortion type. Since the reference image is inaccessible in many applications, the BIQA is quite desirable in this context. In our method, a new feature fusion scheme is first introduced by combining an image's statistical information from multiple domains (i.e., discrete cosine transform, wavelet, and spatial domains) and multiple color channels (i.e., Y, Cb, and Cr). Then, the predicted image quality is generated from a nonparametric model, which is referred to as the label transfer (LT). Based on the assumption that similar images share similar perceptual qualities, we implement the LT with an image retrieval procedure, where a query image's KNNs are searched for from some annotated images. The weighted average of the KNN labels (e.g., difference mean opinion score or mean opinion score) is used as the predicted quality score. The proposed method is straightforward and computationally appealing. Experimental results on three publicly available databases (i.e., LIVE II, TID2008, and CSIQ) show that the proposed method is highly consistent with human perception and outperforms many representative BIQA metrics.**

***Index Terms*— Blind image quality assessment (BIQA), label transfer (LT), multichannel features fusion.**

I. Introduction

DIGITAL images are popular in visual communication, entertainment, and social networks. In these fields, an efficient image quality assessment (IQA) algorithm [1], [2] is crucial to evaluate, control, and enhance the perceptual image quality. Recently, many objective IQA approaches have been developed. Based on the availability of a reference image, these methods are usually classified into three types: 1) full reference (FR); 2) reduced reference (RR); and 3) no reference/blind (NR). The FR-IQA metrics require full access to the undistorted image and the RR-IQA methods also need partial information from the reference image. By incorporating the spatial domain information, transformation domain information [3], [4] and saliency information [5], [6] into the IQA, many well-established FR [7]–[11] and RR [12], [13] metrics have captured human perception well. However, in many practical applications, the reference image is unavailable (e.g., image restoration, etc.), which limits the application fields of the FR and RR IQA algorithms. In contrast, the NR/Blind-IQA method does not need the information of the reference image, which is appealing and quite challenging.

Most existing BIQA approaches are composed of two modules.

1) *Quality-Aware Feature Extraction:* This module generates an efficient image representation to capture the perceptual quality variation caused by the distortion. As discussed in [2] and [14]–[17], many BIQA methods focus on describing an image based on its natural scene statistics (NSS) from the single color channel (i.e., the grayscale map). These NSS are extracted from different domains, e.g., BLINDS-II [15] focuses on the DCT domain, distortion identification-based image verity and integrity evaluation (DIIVINE) [16] works on the wavelet domain, and natural image quality evaluator (NIQE) [17] is executed on the spatial domain.
2) *Prediction Model Learning:* This module is mainly used to map the image features to the subjective quality scores. In many BIQA algorithms [15], [16], [18]–[20], the learning-based regression models are widely used, such as the support vector regression (SVR) [21] and the general regression neural network (GRNN) [22].

Although the aforementioned methods have achieved promising results, many important characteristics of the visual perception are still underutilized.

1) In the feature extraction module, many existing methods extract the single-domain features, which are not sufficient to simulate the complex visual perception mechanism. The multidomain information in both the

Manuscript received March 17, 2014; revised July 25, 2014, December 4, 2014, and February 5, 2015; accepted March 2, 2015. Date of publication March 13, 2015; date of current version March 3, 2016. This work was supported in part by the National Basic Research Program (973 Program) of China under Grant 2015CB351804, in part by the National Natural Science Foundation of China under Grant 61271289, and in part by The program for Science and Technology Innovative Research Team for Young Scholars in Sichuan Province, China, under Grant 2014TD0006. This paper was recommended by Associate Editor Y. Wang.

Q. Wu, H. Li, F. Meng, B. Luo, and C. Huang are with School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: wqb.uestc@gmail.com; corresponding e-mail hlli@uestc.edu.cn; fmmeng@uestc.edu.cn; mathild1987@163.com; huangchao_uestc@aliyun.com).

K. N. Ngan is with the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, and also with the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 610051, China (e-mail: knngan@ee.cuhk.edu.hk).

B. Zeng is with the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 610051, China, and also with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong (e-mail: eezeng@uestc.edu.cn).

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/TCSVT.2015.2412773

spatial and spatial-frequency domains is necessary to precisely represent an image in the visual cortex [23], [24]. Moreover, the multichannel color information is rarely discussed for the BIQA task. As discussed in [25], we know that the trichromacy is an important underlying property of the human vision system.

2) In the quality prediction module, there are also two common issues. First, the learning-based methods are dataset dependent. When the training samples are changed, the model parameters need to be retrained. Second, these regression models (e.g., SVR and GRNN) usually work like the black box mapping, which cannot provide an intuitive visual perception interpretation for the BIQA.

To address the problems mentioned above, we propose a novel BIQA algorithm that is an extension of our previous work [26]. In particular, we introduce the multidomain/channel information to capture the hierarchical and trichromatic properties which are lost in the single-domain/channel features. Meanwhile, a label transfer (LT) method is proposed to intuitively simulate the visual memory retrieval process in the BIQA. In comparison with [26], the visual perception properties behind each proposed feature are further explored here. More extensive experiments and application instances are added to evaluate the proposed method. Meanwhile, the computational complexity of our method is investigated as well.

In view of the superiority of capturing different NSS characteristics under each distortion type, we follow the two-step scheme in [16] and [19] and implement the BIQA with the distortion type classification and LT (TCLT). Inspired by the human perception properties, many new elements are introduced into the proposed method, i.e.,

1) Multiple-domain features are introduced to simulate the hierarchical structure of the visual cortex perception [27], [28], which combine the DCT, wavelet, and spatial domain information to compensate for the visual information lost in the single domain.
2) In view of the trichromatic property of human color vision [25], a multichannel fusion scheme is developed by combining the NSS information from all of the YCbCr color channels.
3) Mittal *et al.* [17] proposed a novel completely blind metric that estimates a query image's quality by measuring its difference with the pristine images in terms of the multivariate Gaussian model-based NSS feature. Inspired by this paper, we develop an LT method to transfer the DMOS labels from some annotated samples to the query image. Based on the assumption that the images with similar quality-aware features share similar perceptual qualities, we utilize the feature distance to search for the query image's KNNs and compute specific weights for each query image. Then, the weighted average of the KNN's DMOS labels is used as the predicted perceptual quality.

The framework of our proposed method is shown in Fig. 1. First, an support vector machine (SVM) classifier is used to identify the query image's probabilities of belonging to each distortion type, which are denoted by p_1 to p_M. Second, the

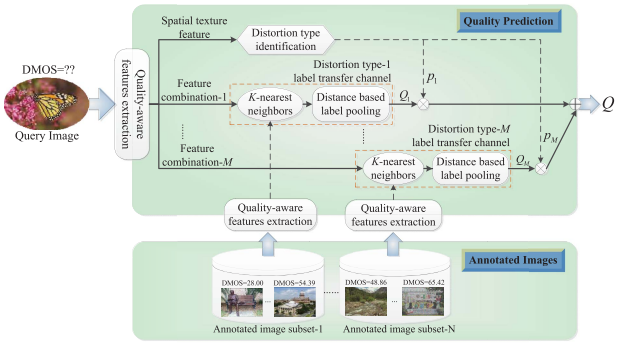


Fig. 1. Framework of our proposed TCLT method.

distortion-specific LT is implemented in each annotated image subset, whose samples share the same distortion type. The predicted qualities in each LT channel are denoted by Q_1 to Q_M. Finally, the weighted average of Q_1 to Q_M is used as our perceptual quality score Q. Extensive experiments on the LIVE II [29], TID2008 [30], and CSIQ [31] databases show that the proposed method is remarkably consistent with human perception and outperforms many state-of-the-art BIQA metrics.

The remainder of this paper is organized as follows. Section II describes the proposed multichannel fusion features. Section III presents the LT-based quality prediction model. Experimental results are shown in Section IV. Then, an image auto-denoising application is discussed in Section V. Finally, the conclusion is given in Section VI.

II. Multichannel Fusion Features

Our crucial idea is to match perceptually similar images. Thanks to the developments of neural science and visual cognition theories, plenty of neurophysiological evidence has been found to reveal the visual perception process.

It is verified that the visual perception system is highly hierarchical, i.e., the local low-level image features are extracted in the neurons of the early visual area and the more complex features will be produced in the higher visual areas. Here, we investigate the properties of areas V1 and V2, which are highly correlated with the visual description in the primate neocortex [27], [28]. Then, the multichannel fusion features in multiple domains are introduced to simulate this perception structure.

A. Quality-Aware Features for Visual Area V1

As described in [27] and [28], V1 is the first visual perception cortical area, which is sensitive to simple local features, e.g., edges, bars, and spatial frequency. Correspondingly, the DCT is a practical tool to represent the local patch with its responses to some orthogonal basis functions, which capture different spatial frequency and local structures [32], [33].

It is noted that human perception presents different sensitivities for different spatial frequencies [27], [33]. Thus, we focus on investigating the NSS of the normalized frequency band coefficient c_i in each band. Let $f_i(x, y)$ denote an element in

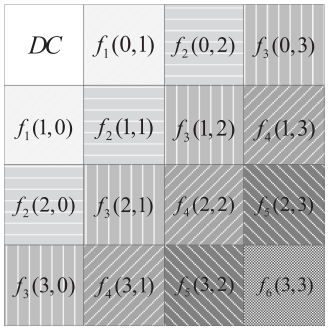


Fig. 2. Frequency bands in the 4×4 DCT domain.

TABLE I
COMPARISON BETWEEN THE PROPOSED DCT FEATURES AND THE OTHER METHODS

	Full-band	Band-wise	Multi-channel
[14]	Mean value, Kurtosis	Renyi-Entropy	×
[15]	GGD shape parameter, Variance/Mean	Energy Difference, Multi-orientation Variance/Mean	×
Proposed	Skewness	Shannon-Entropy, Difference Shannon-Entropy	✓

the xth row, yth column of the DCT coefficient matrix and its frequency band is i. Then, we can get

$$c_i = \sqrt{\frac{1}{N_i} \sum_{x,y \in U_i} f_i^2(x,y)}$$
$$U_i = \{x, y | x + y = i, 0 \le x < W, 0 \le y < W\} \quad (1)$$

where N_i is the number of DCT coefficients in the ith frequency band, and W is the size of the coefficient matrix. An instance of the frequency band locations for a 4×4 block is illustrated in Fig. 2, where the coefficients in the same frequency band are labeled with the same intensity and texture.

Here, we try to develop quality-aware indexes from the DCT domain. Similar works can be found in [14] and [15]. BLINDS [14] describes the local contrast, zero coefficient peakness, and the directional information loss in the DCT domain. BLINDS-II [15] extracts the DCT coefficients' shape parameter, the frequency variation, and the relative distribution between the higher and the lower bands.

Since human vision tends to represent images with minimal redundancy [34], we explore the quality-aware DCT features from the coding-related cues, which is different from previous works. As discussed in [35], NSS mainly concentrates on the intra-block and inter-block correlations in the DCT domain. Here, we develop three DCT domain indices to capture these correlation properties. A detailed comparison between the proposed DCT features and [14] and [15] is shown in Table I, where the full-band denotes the statistics on all DCT coefficients and the band-wise denotes the statistics in each frequency band. In the full-band, we use only the skewness to describe the intra-block correlation and [14], [15] introduce more information, such as, mean, variance, kurtosis, and generalized Gaussian distribution shape parameter. Band-wise, we compute two kinds of Shannon entropies on the normalized frequency band coefficients, which does not consider the correlation in different orientations like [15]. In addition, as shown in Table I, only our DCT features introduce the multichannel color information, which is not considered for both [14] and [15].

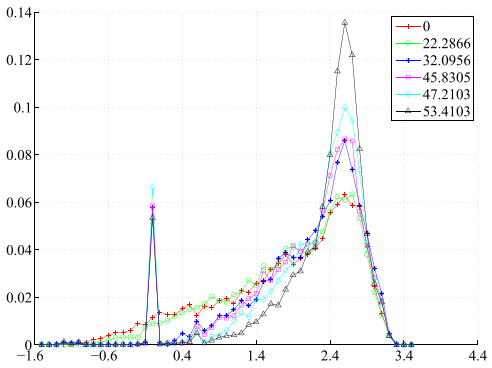


Fig. 3. Distributions of the intra-block skewness for the images with different DMOS. The x-axis indicates the available skewness values, and the y-axis indicates the distribution probability of each skewness value. Legend: the DMOS value of each image.

1) Intra-Block DCT Feature: In the DCT domain, a well-known feature for the natural image is its energy decay property as the frequency increases in the intra-block [35]. When the annoying distortion is present, the decay rate of the frequency band coefficients' distribution will change accordingly. Here, the distribution of the intra-block skewness [36] is computed to measure this statistical variation.

Let $\mathcal{C}^j$ denote the frequency band coefficient set in the jth block, where $\mathcal{C}^j = \{c_1^j, \ldots, c_{N_c}^j\}$ and N_c is the number of the frequency bands in a block. In our experiment, the DCT is implemented on each nonoverlapped 8×8 block, which makes N_c up to 14. Then, the jth block's skewness s^j is

$$s^j = \frac{E(\mathcal{C}^j - \mu(\mathcal{C}^j))^3}{\sigma^3(\mathcal{C}^j)} \quad (2)$$

where $\mu(\cdot)$ is the mean value operator, $\sigma(\cdot)$ is the standard deviation operator, and $E(\cdot)$ is the expectation operator.

To obtain the global descriptor, we compute the intra-block skewness distribution across all blocks. Let $\mathcal{S}$ denote the intra-block skewness set, where $\mathcal{S} = \{s^1, \ldots, s^{N_b}\}$ and N_b is the blocks' number. Then, its margin distribution $P(\mathcal{S})$ is

$$P(\mathcal{S}) = norm(hist(\mathcal{S})) \quad (3)$$

where $hist(\cdot)$ is the histogram operator and $norm(\cdot)$ represents the $l1$-normalization, i.e., $norm(x) = x/\|x\|_1$. Here, the dimension of $P(\mathcal{S})$ is set to 51 based on our experimental study, which could achieve good-quality prediction accuracy.

To illustrate the distortion impact for the intra-block skewness, an instance is shown in Fig. 3, where the curves with different colors denote the $P(\mathcal{S})$ extracted from the luma components of the image *monarch* and its five JPEG2000 (JP2K) versions. A higher skewness indicates a more intensive mass distribution on the left of the mean value, which is induced by a higher decay rate. It is clear that the perceptual quality degradation could be efficiently captured by $\mathcal{P}(\mathcal{S})$, where different image qualities correspond to different distributions. For the severely distorted images labeled by larger DMOS,

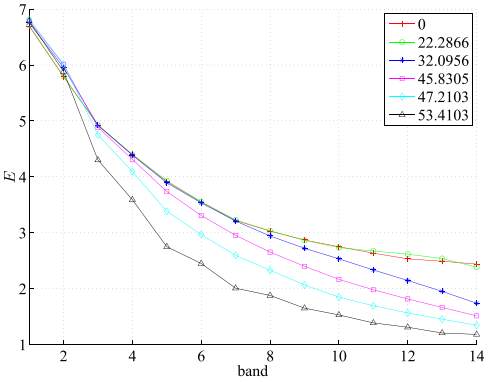


Fig. 4. Distribution of the DCT entropy feature versus the frequency band. The x-axis indicates the frequency band in the DCT domain, and the y-axis indicates the corresponding frequency band entropy. Legend: the DMOS value of each image.

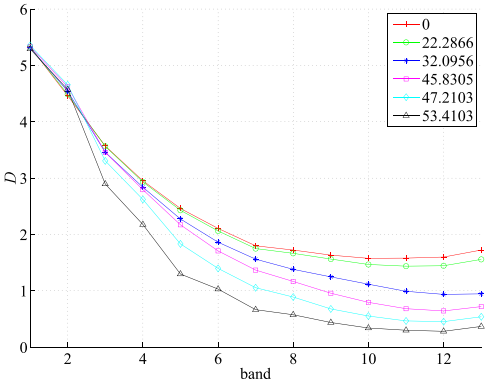


Fig. 5. Distribution of the DCT inter-band difference entropy feature versus the frequency band. The x-axis indicates the frequency band in the DCT domain, and the y-axis indicates the corresponding inter-band difference entropy. Legend: the DMOS value of each image.

their $\mathcal{P}(\mathcal{S})$ would be more focused on higher values. For the slightly distorted images with smaller DMOS, their $\mathcal{P}(\mathcal{S})$ are more smooth and spread on many low-value bins. Since the JP2K introduces a high frequency loss, it accelerates energy decay for the intra-block DCT coefficients. We can measure two images' perceptual similarity by computing their $P(\mathcal{S})$ feature distance, where more similar images have a smaller distance.

2) Inter-Block DCT Feature: It is verified that human vision tends to present natural scenes with minimal redundancy [34], [37]. The image distortions often modify the amount of image information by smoothing local structure or adding random noise. To measure this variation, two entropy-based indices are introduced here.

First, let $\mathcal{C}_i$ denote the coefficient set of the ith frequency band across all blocks, where $\mathcal{C}_i = \{c_i^1, \ldots, c_i^{N_b}\}$. Then, the ith band's Shannon entropy [38] e_i is defined as

$$e_i = E[-\log_2(P(\mathcal{C}_i))] \tag{4}$$

where $P(\mathcal{C}_i) = norm(hist(\mathcal{C}_i))$. In our experiment, we quantize the $\mathcal{C}_i$ into 500 bins in computing its histogram.

The global frequency band entropy feature is represented as

$$E = [e_1, e_2, \ldots, e_{N_c}]. \tag{5}$$

An instance of the distortion impact for E is illustrated in Fig. 4, where different curves correspond to the E extracted from the luma components of the image *monarch* and its five JP2K versions. Similar to the intra-block energy compaction property [39], we can find an entropy decay trend as the DCT frequency band increases in the inter-block statistics. This is because the natural images have higher responses to the low-frequency DCT basis functions. When the local image structures vary from one block to another, the low-frequency coefficients exhibit a larger dynamic range and decentralized distribution. In contrast, the high-frequency coefficients are usually very small and gather around zero. When the JP2K compression is introduced, there are more high-frequency coefficients quantized to zero, which further accelerates the entropy decay, as shown in Fig. 4.

Second, to pop out the distortion effect under diverse image contents [40], we further compute the entropy for the difference g_i of the neighboring frequency bands' coefficients

$$g_i = c_i - c_{i+1}. \tag{6}$$

Let $\mathcal{G}_i$ denote the difference set between the ith and $(i+1)$th frequency bands' coefficients across all blocks, where $\mathcal{G}_i = \{g_i^1, \ldots, g_i^{N_b}\}$. The entropy of $\mathcal{G}_i$ can be represented as

$$d_i = E[-\log_2(P(\mathcal{G}_i))] \tag{7}$$

where $P(\mathcal{G}_i) = norm(hist(\mathcal{G}_i))$. Similar to (4), the histogram bins of $hist(\mathcal{G}_i)$ are also set to 500.

By combining the inter-band difference entropy of all frequency bands, we can obtain the second inter-block feature

$$D = [d_1, d_2, \ldots, d_{N_c-1}]. \tag{8}$$

In Fig. 5, we show an instance to illustrate the distortion impact on feature D, where different curves denote the D extracted from the luma components of the image *monarch* and its five JP2K versions. It can be seen that all curves associated with different DMOS show different distributions. Meanwhile, a larger DMOS shows a greater difference with respect to the undistorted image. Similar to E, the feature D also presents the energy decay trend as the frequency increases, which is caused by the nonlinear amplitude decay of the spatial frequency responses for the natural images [37]. Since the low-frequency bands have a higher decay rate, the dynamic range of their difference is larger relative to the high-frequency bands. When JP2K compression is applied, the small coefficients in the high-frequency bands are more easily eliminated, which enlarges the difference between the high-frequency and low-frequency bands and increases the decay rate.

B. Quality-Aware Features for Visual Area V2

As discussed in [41], visual area V2 contains two features. One focuses on capturing the localized, oriented, and bandpass information like area V1. The other one represents a more complex position and scale invariance and is sensitive to

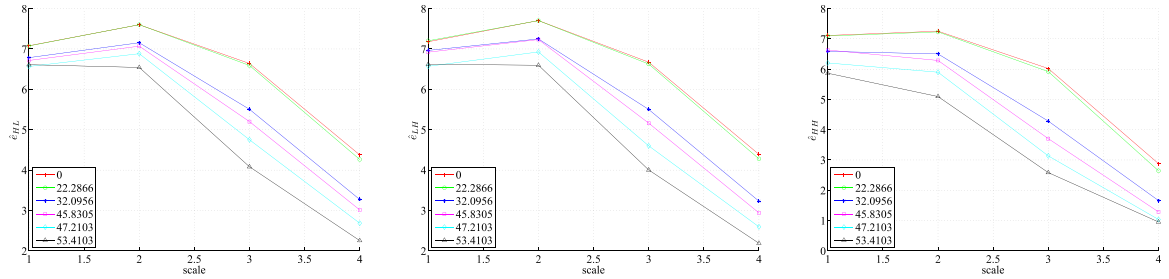


Fig. 6. Distributions of wavelet entropy under different sub-bands. The x-axis indicates the scale in each sub-band and the y-axis indicates the entropy of the wavelet coefficients. Legend: the DMOS of each image. (a) HL. (b) LH. (c) HH.

the shape and texture information [42]. To describe these properties, we develop two statistical indices in the wavelet domain and spatial domain, respectively.

1) Multiscale and Multidirection Information in the Wavelet Domain: Due to the inherent multiscale and multidirection properties, wavelet transform is widely used to simulate the visual cortex property [16]. Wavelet decomposition is usually implemented in a multiscale steerable pyramid structure along the horizontal, vertical, and diagonal directions, which are denoted by HL, LH, and HH, respectively. Here, the wavelet decomposition scale is set to 4 based on our experimental study, which could achieve good-quality prediction accuracy.

In the wavelet domain, there are two important statistical properties for natural images, i.e., exponential decay and self-similarity across all scales [43], [44]. When distortion is introduced, deviation will arise in these statistics. Here, we develop two statistical indices to describe these properties.

First, we employ the entropy of each sub-band to quantitatively measure the exponential decay property. Here, we decompose each channel of a color image into L scales with the wavelet transform. Let $\hat{e}_{k,l}$ denote the entropy of a sub-band in the kth direction and 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

$$\hat{e}_{k,l} = E[-\log_2(P(\mathcal{X}_{k,l}))] \tag{9}$$

where $\mathcal{X}_{k,l}$ denotes the wavelet coefficient set in the kth direction and the lth scale and $P(\mathcal{X}_{k,l}) = norm(hist(\mathcal{X}_{k,l}))$. The bin number of $hist(\mathcal{X}_{k,l})$ is set to 800 based on our experimental study, which could achieve good-quality prediction accuracy.

To capture the exponential decay properties under different directions, we collect three directional sub-band entropies

$$\begin{aligned}\hat{e}_{\mathrm{HL}} &= [\hat{e}_{1,1}, \hat{e}_{1,2}, \ldots, \hat{e}_{1,L}] \\ \hat{e}_{\mathrm{LH}} &= [\hat{e}_{2,1}, \hat{e}_{2,2}, \ldots, \hat{e}_{2,L}] \\ \hat{e}_{\mathrm{HH}} &= [\hat{e}_{3,1}, \hat{e}_{3,2}, \ldots, \hat{e}_{3,L}]\end{aligned} \tag{10}$$

where $\hat{E} = [\hat{e}_{\mathrm{HL}}, \hat{e}_{\mathrm{LH}}, \hat{e}_{\mathrm{HH}}]$ is defined as the overall wavelet entropy feature and its dimension is $3 \times L$.

Fig. 6 shows the instances of the distortion impact on $\hat{e}_{\mathrm{HL}}$, $\hat{e}_{\mathrm{LH}}$, and $\hat{e}_{\mathrm{HH}}$, where different curves correspond to the wavelet entropy features extracted from the luma components of the image *monarch* and its five JP2K versions. It is clear that the curves associated with different perceptual qualities separate from each other. The curves associated with larger DMOS are farther from the undistorted image's curve. In all three directions, the sub-band entropies present a regular decay trend from the coarse scale to the finer one. When JP2K compression is present, the low-value coefficients on the finer scales are more easily quantized to zero, which causes a higher entropy decay rate, as shown in Fig. 6.

Second, the self-similarity property indicates that the wavelet coefficients are strongly correlated across all scales, which is usually referred to as inter-sub-band correlation [45]. Moorthy and Bovik [16] try to measure this property with the structural similarities among different wavelet sub-bands, where the spatial information is considered in locating the comparison windows. In this paper, we consider only the difference of the frequency distribution across the neighboring scales and the Kullback–Leibler divergence (KLD) [38] is employed to measure it. Let $\hat{d}_{k,l}$ denote the neighboring sub-bands' KLD between the lth and the $(l+1)$th scale along the kth direction. Then, its definition can be given by

$$\hat{d}_{k,l} = E_{P(\mathcal{X}_{k,l+1})}\left[\log_2\left(\frac{P(\mathcal{X}_{k,l+1})}{P(\mathcal{X}_{k,l})}\right)\right] \tag{11}$$

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

Similar to our entropy features, the KLD features are also computed in three individual directions

$$\begin{aligned}\hat{d}_{\mathrm{HL}} &= [\hat{d}_{1,1}, \hat{d}_{1,2}, \ldots, \hat{d}_{1,L-1}] \\ \hat{d}_{\mathrm{LH}} &= [\hat{d}_{2,1}, \hat{d}_{2,2}, \ldots, \hat{d}_{2,L-1}] \\ \hat{d}_{\mathrm{HH}} &= [\hat{d}_{3,1}, \hat{d}_{3,2}, \ldots, \hat{d}_{3,L-1}]\end{aligned} \tag{12}$$

where $\hat{D} = [\hat{d}_{\mathrm{HL}}, \hat{d}_{\mathrm{LH}}, \hat{d}_{\mathrm{HH}}]$ is defined as the overall wavelet inter-sub-band KLD feature and its dimension is $3 \times (L-1)$.

The instances of the distortion impact on $\hat{d}_{\mathrm{HL}}$, $\hat{d}_{\mathrm{LH}}$, and $\hat{d}_{\mathrm{HH}}$ are shown in Fig. 7, where different curves denote the inter-sub-band KLD features extracted from the luma components of the image *monarch* and its five JP2K versions. In all directions, the curves associated with different DMOS present different distributions. The seriously distorted image show higher inter-sub-band KLD than the clear image in the mid-scale. However, an opposite case arises in the fine-scale. It is noted that the mid-scale $\hat{d}_{k,l}$ measures the inter-sub-band

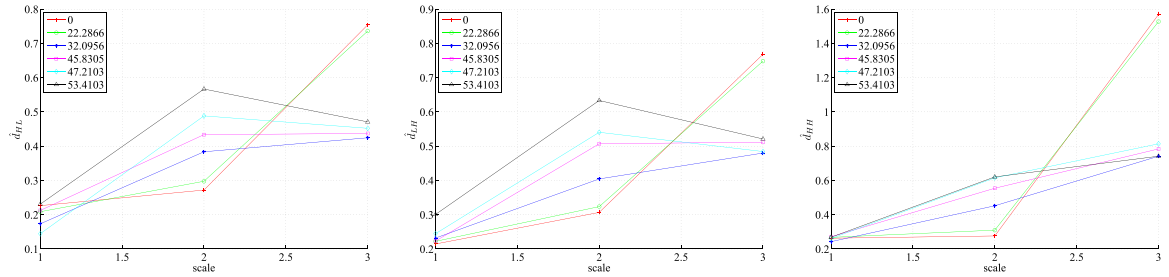


Fig. 7. Distributions of wavelet KLD under different sub-bands. The x-axis indicates the scale in each sub-band and the y-axis indicates the KLD of two neighboring wavelet sub-bands. Legend: the DMOS value of each image. (a) HL. (b) LH. (c) HH.

correlation between the coarse-scale and the fine-scale. Since the magnitudes of the coarse-scale coefficients are higher than those of the fine-scale coefficients, the information loss caused by JP2K is imbalanced in the two scales, which increases the inter-sub-band difference. In contrast, the fine-scale $\hat{d}_{k,l}$ corresponds to two sub-bands in the finer scales. JP2K tends to suppress the low-value coefficients, which makes the coefficient distributions of these two scales more consistent.

2) Texture Information in the Spatial Domain: As discussed in [46] and [47], most cells in area V2 possess the cue-invariant responses for the texture. To capture this property, we use the local binary pattern (LBP) [48] descriptor, which has gray-scale and rotation invariance. As discussed in [48], we know that LBP can obtain a good texture classification performance with 16 neighbors in the radius of 2. A larger neighbor and the radius may slightly increase its classification accuracy. However, the feature dimension would become higher, which increases the complexity in computing the feature distance. Thus, we set the radius to 2 and use 16 neighbors in computing the LBP^{ri} feature, whose dimension is 4116.

C. Multichannel Fusion for Trichromatic Property

We introduce the multichannel fusion features based on two factors. First, human vision has an inherent trichromatic property [25]. Second, some image distortions may present different degradation degrees in each color channel. For example, JPEG and JP2K use different quantization and entropy coding settings for the luma and chroma channels.

Here, we use the YCbCr color space for its superiority in matching human perception and the low complexity in color space transformation [49]. Since the Cb and Cr components present strong correlation [50], the same features are extracted from them. The multichannel fusion features can be obtained by combining the NSS from all color channels.

Here, the DCT domain multichannel fusion features are

$$\begin{aligned}\mathcal{P} &= [P^{\text{Y}}(\mathcal{S}), P^{\text{Cb}}(\mathcal{S}), P^{\text{Cr}}(\mathcal{S})]\\ \mathcal{E} &= [E^{\text{Y}}, E^{\text{Cb}}, E^{\text{Cr}}]\\ \mathcal{D} &= [D^{\text{Y}}, D^{\text{Cb}}, D^{\text{Cr}}]\end{aligned} \tag{13}$$

where the dimensions of $\mathcal{P}$, $\mathcal{E}$, and $\mathcal{D}$ increase to 3×51, 3×14, and 3×13, respectively.

TABLE II
MEDIAN SROCC ACROSS 100 TRAIN-TEST TRIALS BASED ON THE NEAREST NEIGHBOR QUALITY ESTIMATION

	JP2K	JPEG	WN	Blur	FF	All
$\mathcal{P}$	0.9000	0.9429	0.9000	0.9747	0.8944	0.8742
$\mathcal{E}$	0.8857	0.9276	1.000	0.9000	0.8000	0.8538
$\mathcal{D}$	0.8117	0.9009	1.000	0.9000	0.8000	0.8431
$\hat{\mathcal{E}}$	0.9258	0.8944	1.000	1.000	0.8944	0.8868
$\hat{\mathcal{D}}$	0.8986	0.8286	0.9747	0.9747	0.8721	0.8673
LBP	0.8857	0.9276	1.000	0.9747	0.8944	0.8670

The wavelet domain multichannel features are

$$\begin{aligned}\hat{\mathcal{E}} &= [\hat{E}^{\text{Y}}, \hat{E}^{\text{Cb}}, \hat{E}^{\text{Cr}}]\\ \hat{\mathcal{D}} &= [\hat{D}^{\text{Y}}, \hat{D}^{\text{Cb}}, \hat{D}^{\text{Cr}}]\end{aligned} \tag{14}$$

where the dimensions of $\hat{\mathcal{E}}$ and $\hat{\mathcal{D}}$ increase to $3\times3\times4$ and $3\times3\times3$, respectively.

In the spatial domain, we capture the texture variation caused by distortion. However, most regions in the chroma components are smooth. Since the LBP loses the magnitude of the local contrast, many small noise in the chroma components would change the distribution of LBP even if the human perception is insensitive to these noise. Accordingly, we only extract the LBP feature in the luma component.

D. Validity Analysis

Here, we further give a quantitative analysis to verify the validity of our proposed features, i.e., whether the close samples in our proposed feature space share the similar subjective qualities. We implement a simple nearest neighbor quality prediction on the LIVE II database, where one of 29 reference images and its distorted versions are used as the testing set. The rest of the images are used as the training set.

In the testing stage, we estimate the query image subjective quality with the DMOS of an annotated image, whose feature has the minimum chi-square distance with respect to it. For each proposed feature, we repeat the random nonoverlapped train-test trial 100 times. Finally, the median Spearman's rank ordered correlation coefficients (SROCCs) between the predicted qualities and their ground-truth DMOS across the 100 trials are reported in Table II, where each row corresponds to the performance of one proposed feature.

It can be seen that our proposed features work well in measuring the images' perceptual similarity. Based on a simple

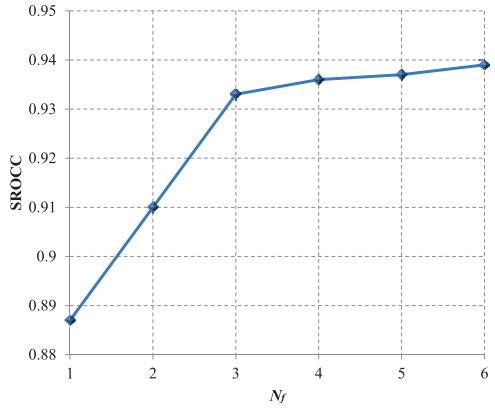


Fig. 8. Plot of the median SROCC versus the number of proposed features.

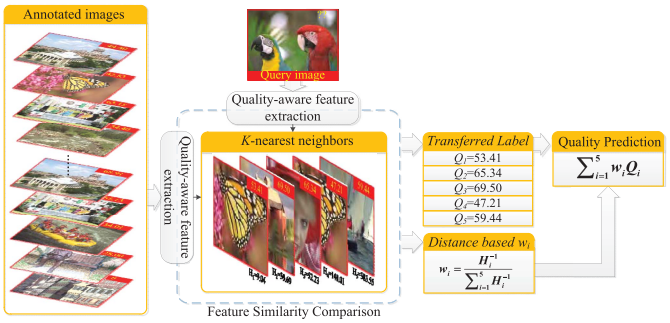


Fig. 9. Diagram of the distortion-specific LT for JP2K. The DMOS label of each annotated image is marked in its top-right corner. H_i denotes the feature distance between the query image and its ith nearest neighbor.

nearest neighbor estimation, we can efficiently predict the test images' quality ranks, where all features' median SROCCs are larger than 0.8 across different distortion types. To verify the validity of our composite features, we investigate the performances of the feature combinations based on the greedy search [51], where the number of composite features N_f can range from 1 to 6. When N_f is larger than one, we use the product of multiple feature distances to measure perceptual similarity. Here, the investigation result is shown in Fig. 8. It can be seen that the median SROCC steadily increases as we add the number of the composite features. That is, our composite features could measure the perceptual similarity more accurately. In addition, since there is no content overlap between the training set and testing set, we can safely conclude that the proposed features could accurately measure the perceptual similarity across different image contents.

III. Prediction Model Based on Label Transfer

As discussed in [52] and [53], the visual memory and perception are highly correlated and share a certain brain mechanism. According to the speculative theoretical framework in [53], the visual system is assumed to be composed of the early visual processing and visual memory modules. Here, the visual memory consists of many natural scene templates and the perception or cognition task can be achieved by retrieving the matched pattern for the query image.

From an application perspective, we simulate this visual memory retrieval with an LT procedure, which transfers the subjective quality labels [e.g., DMOS or mean opinion score (MOS)] from the annotated images to the query image. It is based on the assumption that the images with similar features should share similar perceptual qualities. An instance of the distortion-specific LT for JP2K is shown in Fig. 9. Here, we first search for the query image's KNNs from the annotated set. There are two outputs in this stage, i.e., the transferred DMOS labels Q_i and the distance-based weight w_i. Then, we can obtain the predicted quality by integrating Q_i and w_i. This LT method could adaptively update w_i for each test image, which works like piecewise regression and improves the robustness of the regression model [54]. Following the two-step scheme in [16], we design the distortion-specific LT channel for each distortion type.

First, we divide the annotated images into distortion-specific subsets, which are used for different LT channels, respectively. We identify the query image's distortion type with the offline trained SVM classifier [55], whose input is the single-channel LBP feature and the outputs are the query image's probabilities of belonging to each distortion type. To prevent overfitting, cross validation is employed to determine the classifier's parameters [56], [57].

Second, the distortion-specific LT is separately implemented in each LT channel. We search for the query image's KNNs in the distortion-specific image subset and the KNN's labels are transferred to the query image. Then, the weighted average of the received labels is used as the predicted quality score in each LT channel.

Let $\mathcal{F}_i$ denote the feature vector set of the ith query image, where $\mathcal{F}_i = \{F_i^1, \ldots, F_i^{N_F}\}$ and N_F is the number of the feature vectors. Let $\widetilde{\mathcal{F}}_j^m$ denote the feature vector set of the jth reference sample in the annotated image subset with the mth distortion type, where $\widetilde{\mathcal{F}}_j^m = \{\widetilde{F}_j^{m,1}, \ldots, \widetilde{F}_j^{m,N_F}\}$. As discussed in [58] and [59], the product combination from multiple smaller classifiers could achieve the optimal classification performance for the independent training data. Thus, we use the product fusion to combine all feature similarities. We define the distance $H_{m,j}$ between $\mathcal{F}_i$ and $\widetilde{\mathcal{F}}_j^m$ as the product of the chi-square distances for each pair of the feature vectors

$$H_{m,j} = \prod_{k=1}^{N_F} h\left(F_i^k, \widetilde{F}_j^{m,k}\right)$$
$$h\left(F_i^k, \widetilde{F}_j^{m,k}\right) = \sum_{t=1}^{N_k} \frac{\left(F_i^k(t) - \widetilde{F}_j^{m,k}(t)\right)^2}{F_i^k(t) + \widetilde{F}_j^{m,k}(t)} \quad (15)$$

where $F_i^k(t)$ and $\widetilde{F}_j^{m,k}(t)$ denote the tth element of F_i^k and $\widetilde{F}_j^{m,k}$, N_k is the dimension of the kth feature vector.

Based on our experimental study, we select the top five nearest neighbors of the query image from the current annotated image subset in terms of $H_{m,j}$, which achieves good-quality prediction accuracy at low complexity. Then, a distance-based weighting scheme is used for the KNN's labels (i.e., DMOS) to predict the subjective quality in current LT channel. Here, a larger weight is assigned to the label which has a smaller feature distance. The normalized weight $w_{m,j}$ for the jth

TABLE III
PERCEPTION CONSISTENCY PERFORMANCE OF EACH FEATURE

	JP2K	JPEG	WN	Blur	FF	All	Dimension
$\mathcal{P}$	0.8663	**0.8959**	0.9451	0.8988	0.8178	**0.8957**	153
$\mathcal{E}$	0.8398	0.8708	0.9789	0.9179	**0.8365**	0.8929	42
$\mathcal{D}$	0.8116	0.8785	**0.9818**	0.8997	0.8329	0.8830	39
$\hat{\mathcal{E}}$	0.8429	0.8471	0.9804	**0.9299**	0.8289	0.8791	36
$\hat{\mathcal{D}}$	0.8524	0.8346	0.9135	0.9250	0.8440	0.8748	27
LBP	**0.8680**	0.8569	0.9680	0.9239	0.8118	0.8878	4116

selected neighbor image can be defined as

$$w_{m,j} = H_{m,j}^{-1} \Big/ \sum_{j=1}^{5} H_{m,j}^{-1}. \tag{16}$$

The predicted quality of the mth LT channel is

$$Q_m = \sum_{j=1}^{5} w_{m,j} \cdot \mathrm{DMOS}_j. \tag{17}$$

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

$$Q = \sum_{m=1}^{M} p_m \cdot Q_m \tag{18}$$

where p_m is the query image's probability of belonging to the mth distortion type, and M is the number of the distortion types.

IV. EXPERIMENTAL RESULTS

A. Protocol

We test the proposed method on the LIVE II [29], TID2008 [30], and CSIQ [31] databases. The MATLAB code of our method has been released online.[1] The consistency experiment is conducted on the LIVE II database, which contains five distortion types, i.e., JP2K, JPEG, additive white noise (WN), Gaussian blur (Blur), and fast fading (FF). Similar to [14]–[16], a cross validation is implemented by randomly splitting the LIVE II database into two nonoverlapped sets. We use 23 of 29 reference images and their associated distorted images as the training set. The remaining images make up the test set.

The training set is used to learn the distortion-type classifier and construct the annotated image set. Here, we conduct the random splitting evaluation 100 times. The median values of the indices across the 100 trials are used for verification. Four measures are adopted to compare different BIQA approaches, i.e., the Pearson's linear correlation coefficient (PLCC), the SROCC, the root-mean-square error (RMSE), and mean absolute error (MAE) between the predicted quality Q and the DMOS.

B. Feature Analysis

Since the discriminatory abilities of different features vary for different distortion types, we first investigate each feature's performance. Each candidate feature is separately used to compute the similarity between the query image and the annotated reference sample across all distortion types. Here, we use only the data from the training set of the LIVE II database. Similar to Section IV-A, we divide these training images into the content nonoverlapped annotation set and validation set, which contains 60% and 20% samples, respectively. The median SROCCs of the validation set over 100 train-test trials are reported in Table III, where the optimal results in each row are labeled in boldface and the dimension of each proposed feature is listed in the last column.

It is seen that the SROCC values of the proposed features are all larger than 0.8 under different distortion types. Meanwhile, no single feature performs best across all distortion types. Thus, it is natural to select different feature combinations for each distortion-specific LT channel.

Let $\rho_{m,n}^k$ denote the SROCC obtained from the LT of the kth feature, where the query image belongs to the mth distortion type and the annotated image subset possesses the nth distortion type. Let σ_n^k denote the standard derivation of the SROCC for the kth feature in the nth distortion-specific LT channel. Then, we can get

$$\begin{aligned} \left(\sigma_n^k\right)^2 &= \frac{1}{M} \sum_{m=1}^{M} \left(\rho_{m,n}^k - \mu_n^k\right)^2 \\ \mu_n^k &= \frac{1}{M} \sum_{m=1}^{M} \rho_{m,n}^k. \end{aligned} \tag{19}$$

By considering the classification accuracy, we represent the kth feature contribution for the query images with the mth distortion type as I_m^k

$$\begin{aligned} I_m^k &= \sum_{n=1}^{M} \frac{\rho_{m,n}^k \cdot \mathrm{sgn}(n)}{\sigma_n^k} \\ \mathrm{sgn}(n) &= \begin{cases} a_n & n = m \\ 1 - a_n & n \neq m \end{cases} \end{aligned} \tag{20}$$

where a_n is the nth distortion type's classification accuracy.

TABLE IV
EACH FEATURE'S CONTRIBUTION FOR DIFFERENT DISTORTION-SPECIFIC LT CHANNELS

	DCT			Wavelet		Spatial
	$\mathcal{P}$	$\mathcal{E}$	$\mathcal{D}$	$\hat{\mathcal{E}}$	$\hat{\mathcal{D}}$	LBP
JP2K	**6.5919**	2.1120	2.0308	**5.1754**	**3.1481**	**2.2166**
JPEG	**4.0333**	2.8089	**3.8354**	**8.0938**	**3.0647**	**4.2231**
WN	3.0366	**5.5851**	3.9220	**4.4875**	**5.5007**	**4.2525**
Blur	2.5583	2.9053	**5.0398**	**6.7952**	**3.3566**	**3.5072**
FF	**4.4353**	2.0834	1.9222	**4.2192**	**2.8536**	**3.3435**

TABLE V
FEATURE COMBINATIONS FOR DIFFERENT DISTORTION-SPECIFIC LT CHANNELS

Feature combination ID	Distortion type	Selected features	Dimension
1	JP2K	$\mathcal{P},\hat{\mathcal{E}},\hat{\mathcal{D}}$,LBP	4312
2	JPEG	$\mathcal{P},\mathcal{D},\hat{\mathcal{E}},\hat{\mathcal{D}}$,LBP	4351
3	WN	$\mathcal{E},\hat{\mathcal{E}},\hat{\mathcal{D}}$,LBP	4201
4	Blur	$\mathcal{D},\hat{\mathcal{E}},\hat{\mathcal{D}}$,LBP	4198
5	FF	$\mathcal{P},\hat{\mathcal{E}},\hat{\mathcal{D}}$,LBP	4312

[1] http://ivipc.uestc.edu.cn/wqb/projects/TCLT-release.zip

TABLE VI

MEDIAN PLCC, SROCC, RMSE, AND MAE ACROSS 100 TRAIN-TEST TRIALS ON THE LIVE II IQA DATABASE

Distortion		JP2K				JPEG			
Metric	Type	PLCC	SROCC	RMSE	MAE	PLCC	SROCC	RMSE	MAE
PSNR	FR	0.896	0.890	7.187	5.528	0.860	0.841	8.170	6.380
SSIM	FR	0.937	0.932	5.671	4.433	0.928	0.903	5.947	4.485
VIF	FR	0.962	0.953	4.449	3.445	0.943	0.913	5.321	3.807
pLSA	Blind	0.87	0.85	--	--	0.90	0.88	--	--
BIQI	Blind	0.750	0.736	16.540	--	0.630	0.591	24.580	--
BLINDS	Blind	0.807	0.805	14.780	--	0.597	0.552	25.320	--
DIIVINE	Blind	0.922	0.913	9.660	--	0.921	0.910	12.250	--
BLINDS-II	Blind	**0.963**	**0.951**	--	--	**0.979**	0.942	--	--
BRISQUE	Blind	0.923	0.914	--	--	0.974	**0.965**	--	--
NIQE	Blind	0.937	0.917	--	--	0.956	0.938	--	--
NSS-TS	Blind	0.947	0.931	5.792	7.169	0.933	0.915	6.333	7.912
$\overline{\text{TCLT}}$-Gray	Blind	0.865	0.862	8.654	6.551	0.933	0.910	6.005	4.236
TCLT-Gray	Blind	0.846	0.839	9.067	7.000	0.929	0.912	6.165	4.424
$\overline{\text{TCLT}}$-YCbCr	Blind	0.893	0.889	7.834	6.171	0.948	0.932	5.340	4.023
TCLT-YCbCr	Blind	0.902	0.898	7.721	6.199	0.946	0.923	5.501	4.425
Distortion		WN				Blur			
Metric	Type	PLCC	SROCC	RMSE	MAE	PLCC	SROCC	RMSE	MAE
PSNR	FR	0.986	0.985	2.680	2.164	0.783	0.782	9.772	7.743
SSIM	FR	0.970	0.963	3.916	3.257	0.874	0.894	7.639	5.760
VIF	FR	0.984	0.986	2.851	2.304	0.974	0.973	3.533	2.818
pLSA	Blind	0.87	0.80	--	--	0.88	0.87	--	--
BIQI	Blind	0.968	0.958	6.930	--	0.800	0.778	11.100	--
BLINDS	Blind	0.914	0.890	11.270	--	0.870	0.834	9.080	--
DIIVINE	Blind	0.988	**0.984**	4.310	--	0.923	0.921	7.070	--
BLINDS-II	Blind	0.985	0.978	--	--	0.948	0.944	--	--
BRISQUE	Blind	0.985	0.979	--	--	0.951	**0.951**	--	--
NIQE	Blind	0.977	0.966	--	--	0.953	0.934	--	--
NSS-TS	Blind	0.963	0.971	4.464	6.018	0.950	0.939	5.481	6.863
$\overline{\text{TCLT}}$-Gray	Blind	0.989	0.980	2.571	2.080	0.941	0.928	5.778	4.576
TCLT-Gray	Blind	0.989	0.980	2.679	2.205	**0.955**	0.947	5.085	3.880
$\overline{\text{TCLT}}$-YCbCr	Blind	0.988	0.976	2.638	2.123	0.954	0.941	5.106	4.050
TCLT-YCbCr	Blind	**0.989**	0.979	2.683	2.166	0.954	0.940	5.157	4.057
Distortion		FF				Entire database			
Metric	Type	PLCC	SROCC	RMSE	MAE	PLCC	SROCC	RMSE	MAE
PSNR	FR	0.890	0.890	7.516	5.800	0.824	0.820	9.124	7.325
SSIM	FR	0.943	0.941	5.485	4.297	0.863	0.851	8.126	6.275
VIF	FR	0.962	0.965	4.502	3.547	0.950	0.953	5.024	3.887
pLSA	Blind	0.84	0.77	--	--	0.79	0.80	--	--
BIQI	Blind	0.722	0.700	19.480	--	0.740	0.726	18.360	--
BLINDS	Blind	0.743	0.678	18.620	--	0.680	0.663	20.010	--
DIIVINE	Blind	0.888	0.863	12.930	--	0.917	0.916	10.900	--
BLINDS-II	Blind	**0.944**	0.927	--	--	0.923	0.920	--	--
BRISQUE	Blind	0.903	0.877	--	--	**0.942**	**0.940**	--	--
NIQE	Blind	0.913	0.859	--	--	0.915	0.914	--	--
NSS-TS	Blind	0.942	**0.935**	5.232	7.070	0.926	0.930	5.131	6.803
$\overline{\text{TCLT}}$-Gray	Blind	0.878	0.835	8.362	6.003	0.915	0.910	6.712	4.704
TCLT-Gray	Blind	0.894	0.861	7.842	5.933	0.917	0.916	6.612	4.789
$\overline{\text{TCLT}}$-YCbCr	Blind	0.898	0.872	7.571	5.717	0.927	0.925	6.268	4.494
TCLT-YCbCr	Blind	0.923	0.903	6.964	5.160	0.935	0.934	5.938	4.326

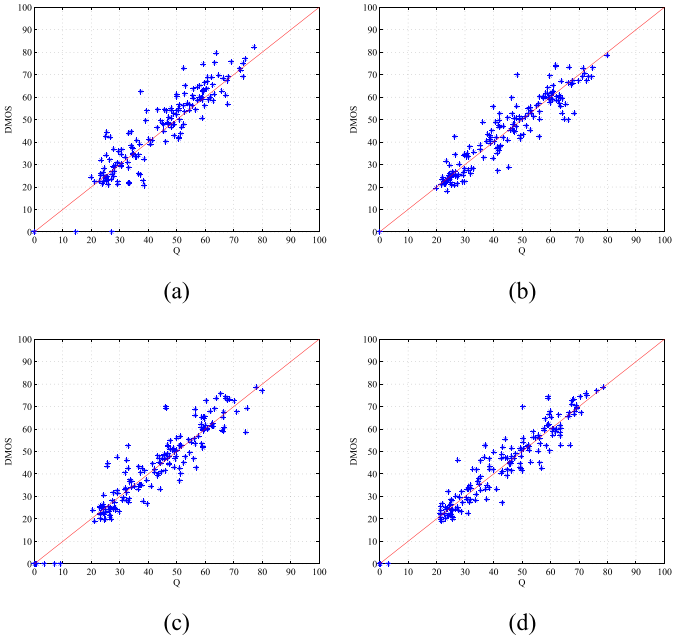


Fig. 10. Scatter plots of the predicted quality index Q versus the DMOS. The x-axis is the predicted quality index Q and the y-axis is the DMOS value. The red line represents the ideal linearly correlated line. (a) $\overline{\text{TCLT}}$-Gray. (b) $\overline{\text{TCLT}}$-YCbCr. (c) TCLT-Gray. (d) TCLT-YCbCr.

From (20), we know that only the feature that produces better correlation performance with DMOS (i.e., higher SROCC) and stronger robustness across different distortion types (i.e., smaller SROCC standard deviation) can bring greater contribution to current LT channel. Table IV shows the I_m^k values, where a higher value in each row means a bigger contribution for the corresponding distortion type. For each distortion type, we select some specific features with the top I_m^k values in each row of Table IV. The number of selected features is determined by cross validation, which achieves the highest median SROCC in the validation set. For clarity, the selected features are labeled in the boldface in Table IV. Meanwhile, we summarize the feature combination information for each distortion-specific LT channel in Table V, where the dimensions of the combined features are listed in the last column.

C. Consistency Experiment

Our method has two characteristics that guarantee high consistency with human perception, i.e., multichannel feature fusion and distortion-specific LT. To separately evaluate them, we implement our method with four combinations: single-channel feature + universal distortion LT (named $\overline{\text{TCLT}}$-Gray); multichannel feature fusion + universal distortion LT (named $\overline{\text{TCLT}}$-YCbCr); single-channel feature + distortion-specific LT (named TCLT-Gray); and multichannel feature fusion + distortion-specific LT (named TCLT-YCbCr). The single-channel features are extracted from the gray image. The universal distortion LT does not identify the distortion type, and evaluates the query image with the weighted average of its KNN's labels searched from all reference images. In measuring two images' similarity, the universal distortion LT would combine all proposed features.

TABLE VII

RESULTS OF THE ONESIDED *t-Test* PERFORMED BETWEEN THE SROCC VALUES OBTAINED FROM FOUR VERSIONS OF THE PROPOSED METHOD. A VALUE OF 1/0/−1 INDICATES THE ROW ALGORITHM IS STATISTICALLY SUPERIOR/EQUIVALENT/INFERIOR TO THE COLUMN ALGORITHM

	$\overline{\text{TCLT}}$-Gray	TCLT-Gray	$\overline{\text{TCLT}}$-YCbCr	TCLT-YCbCr
$\overline{\text{TCLT}}$-Gray	0	-1	-1	-1
TCLT-Gray	1	0	-1	-1
$\overline{\text{TCLT}}$-YCbCr	1	1	0	-1
TCLT-YCbCr	1	1	1	0

Here, we compare our method with some FR-IQA (e.g., peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [7], and visual information fidelity [8]) and BIQA (e.g., blind image quality index [19], probabilistic latent semantic analysis [60], BLINDS [14], DIIVINE [16], BLINDS-II [15], natural image quality evaluator [17], NSS-TS [61], and blind/referenceless image spatial quality evaluator (BRISQUE) [62]) algorithms.

The scatter plots for the four versions of our method are shown in Fig. 10. It can be seen that the predicted quality Q shows a nearly linear relationship with DMOS. For all of $\overline{\text{TCLT}}$-Gray, $\overline{\text{TCLT}}$-YCbCr, TCLT-Gray, and TCLT-YCbCr, the plots are closely distributed around the linearly correlated line DMOS $= Q$. It demonstrates that all of our methods can achieve high consistency with human perception. Meanwhile, some outliers can also be found in Fig. 10, which are far from the linear-correlation line. In Fig. 10(a), the prediction error of the JP2K version of *stream* is up to 25. In Fig. 10(b), the prediction errors for the JP2K and FF versions of *paintedhouse* are 24 and 19, respectively. In Fig. 10(c), the prediction error of the JP2K version of *coinsinfountain* is 22. In Fig. 10(d), the prediction error for the FF version of *ocean* is 18.

The detailed consistency performances are shown in Table VI. The best BIQA metrics' PLCC and SROCC are highlighted in boldface under each distortion type. It can be seen that our predicted qualities are highly consistent with human perception across different distortion types. For Blur, the TCLT-Gray achieves the best PLCC (0.955) and the suboptimal SROCC (0.947) result. For WN, all the $\overline{\text{TCLT}}$-Gray, TCLT-Gray, and TCLT-YCbCr achieve the best performance in terms of PLCC (0.989). Both of the $\overline{\text{TCLT}}$-Gray and TCLT-Gray obtain suboptimal SROCC (0.980) results, which are close to the best DIIVINE result (SROCC $=$ 0.984). For JPEG and FF, the $\overline{\text{TCLT}}$-YCbCr and TCLT-YCbCr achieve the fourth best performance, respectively. For JP2K, the PLCC of TCLT-YCbCr is still more than 0.9.

For the general purpose BIQA, the predicted quality should work well for all distortion types. Here, our TCLT-YCbCr method achieves the second best result (SROCC $=$ 0.934) in the entire database test, which is very close to the state-of-the-art BRISQUE metric (SROCC $=$ 0.940). A onesided *t-test* [63] is also executed on the SROCC of the proposed method and BRISQUE across 100 train-test trails. The reported result shows that the four versions of our proposed method are statistically inferior to BRISQUE in this test.

TABLE VIII

STANDARD DERIVATION OF THE PERFORMANCES OF $\overline{\text{TCLT}}$-GRAY AND $\overline{\text{TCLT}}$-YCbCr ACROSS 100 TRAIN-TEST TRIALS ON THE LIVE II IQA DATABASE

	$\overline{\text{TCLT}}$-Gray				$\overline{\text{TCLT}}$-YCbCr			
	PLCC	SROCC	RMSE	MAE	PLCC	SROCC	RMSE	MAE
JP2K	0.040	0.045	1.251	1.038	0.037	0.042	1.291	1.062
JPEG	0.020	0.032	0.945	0.727	0.017	0.026	0.937	0.631
WN	0.003	0.007	0.364	0.310	0.003	0.008	0.367	0.324
Blur	0.029	0.036	0.949	0.826	0.021	0.026	0.937	0.790
FF	0.085	0.090	2.327	1.465	0.095	0.104	2.767	1.547
All	0.023	0.021	0.866	0.539	0.023	0.023	0.912	0.467

TABLE IX

STANDARD DERIVATION OF THE PERFORMANCES OF TCLT-GRAY AND TCLT-YCbCr ACROSS 100 TRAIN-TEST TRIALS ON THE LIVE II IQA DATABASE

	TCLT-Gray				TCLT-YCbCr			
	PLCC	SROCC	RMSE	MAE	PLCC	SROCC	RMSE	MAE
JP2K	0.046	0.052	1.496	1.162	0.038	0.038	1.329	1.014
JPEG	0.021	0.034	0.948	0.727	0.016	0.027	0.910	0.653
WN	0.003	0.008	0.373	0.346	0.003	0.008	0.385	0.340
Blur	0.019	0.021	0.858	0.709	0.020	0.025	0.919	0.745
FF	0.079	0.074	2.309	1.447	0.060	0.058	2.324	1.443
All	0.023	0.022	0.947	0.554	0.020	0.019	0.874	0.478

TABLE X

MEDIAN CLASSIFICATION ACCURACY (%) ACROSS 100 TRAIN-TEST TRIALS ON THE LIVE II IQA DATABASE

	JP2K	JPEG	WN	Blur	FF	All
DIIVINE	80.00	81.10	100.00	90.00	73.30	83.75
TCLT	85.71	100.00	100.00	96.67	80.00	90.85

TABLE XI

STANDARD DERIVATION OF THE CLASSIFICATION ACCURACY ACROSS 100 TRAIN-TEST TRIALS ON THE LIVE II IQA DATABASE

	JP2K	JPEG	WN	Blur	FF	All
TCLT	8.56	3.38	1.54	2.95	10.91	2.62

In addition, the entire database test verifies that the multichannel feature fusion and distortion-specific LT are both valid in improving the BIQA performance. First, the multichannel fusion features capture more comprehensive visual perception properties than the single-channel features. Thus, it is found that $\overline{\text{TCLT}}$-YCbCr (SROCC $=$ 0.925) outperforms $\overline{\text{TCLT}}$-Gray (SROCC $=$ 0.910) and TCLT-YCbCr (SROCC $=$ 0.934) outperforms TCLT-Gray (SROCC $=$ 0.916). Second, the distortion-specific LT provides a refined reference image subset, which helps the query image to find more accurate references than the universal distortion LT. So it is seen that TCLT-Gray outperforms $\overline{\text{TCLT}}$-Gray and TCLT-YCbCr outperforms $\overline{\text{TCLT}}$-YCbCr. TCLT-YCbCr, which combines both of the aforementioned traits, performs best in all versions of our method. To evaluate the statistical significance of the difference between four versions of our method, we perform the onesided *t-test* [63] on their SROCC across 100 train-test trials. As shown in Table VII, we can find that TCLT-YCbCr still consistently outperforms the other versions of the proposed method.

In Tables VIII and IX, we further show the standard derivations of the four indices for the proposed methods across 100 train-test trials. It can be seen that the standard derivations of the four indices are very small for all the proposed methods. It demonstrates that the performance variations of the proposed methods are insignificant in the 100 trials.

TABLE XII
SROCC BETWEEN THE PREDICTED QUALITY INDEX AND MOS ON THE KNOWN SUBSETS OF THE TID2008 DATABASE

Metric	Type	JP2K	JPEG	WN	Blur	All
PSNR	FR	0.825	0.876	0.918	0.934	0.870
SSIM	FR	0.963	0.935	0.817	0.960	0.902
DIIVINE	Blind	0.924	0.866	**0.851**	0.862	0.889
BLINDS-II	Blind	0.915	0.889	0.696	0.857	0.854
BRISQUE	Blind	0.832	**0.924**	0.829	0.881	**0.896**
TCLT-Gray	Blind	0.880	0.783	0.794	0.870	0.845
TCLT-Gray	Blind	0.863	0.760	0.779	0.849	0.843
TCLT-YCbCr	Blind	**0.925**	0.874	0.802	**0.879**	0.872
TCLT-YCbCr	Blind	0.890	0.895	0.830	0.872	0.877

TABLE XIII
SROCC BETWEEN THE PREDICTED QUALITY INDEX AND DMOS ON THE KNOWN SUBSETS OF THE CSIQ DATABASE

Metric	Type	JP2K	JPEG	WN	Blur	All
PSNR	FR	0.936	0.888	0.936	0.925	0.921
SSIM	FR	0.958	0.944	0.898	0.958	0.926
DIIVINE	Blind	0.830	0.704	0.797	0.871	0.828
BLINDS-II	Blind	0.884	0.881	0.886	0.870	0.873
BRISQUE	Blind	0.857	0.877	0.926	0.874	0.881
TCLT-Gray	Blind	0.868	0.864	0.878	0.874	0.858
TCLT-Gray	Blind	0.876	0.890	0.883	0.895	0.878
TCLT-YCbCr	Blind	0.894	0.898	**0.928**	**0.896**	0.886
TCLT-YCbCr	Blind	**0.896**	**0.907**	0.908	0.882	**0.891**

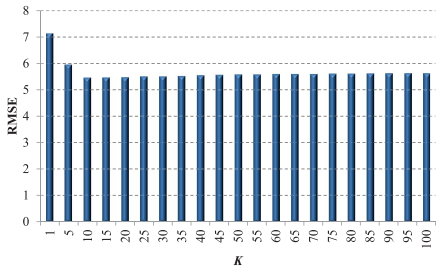


Fig. 11. Median RMSE variation of TCLT-YCbCr under different K.

D. Classification Accuracy

The distortion-specific LT assumes that the reference subset, which has the same distortion type with the query image, improves the accuracy of the LT. Thus, the classification accuracy plays an important role in our TCLT-YCbCr method. Here, we investigate the median accuracies of our distortion type classification across 100 train-test trials. As shown in Table X, it is seen that our classification accuracy is significantly superior to that of DIIVINE for every distortion type and the entire database. Especially for JPEG and WN, our classification accuracies are both up to 100%.

In Table XI, we also show the standard derivations of the classification accuracies across 100 train-test trials. It is seen that the standard derivations of our distortion type classification are very small, which demonstrates that our classification performance varies very slightly in the 100 trials.

E. Database Independency

To verify that the proposed method is independent of the dataset, we further test our method on the subsets of the TID2008 [30] and CSIQ [31] databases. The entire LIVE II database is used as the train set. Four types of known distortions in the train set are selected from TID2008 and CSIQ to construct the test set, i.e., JP2K, JPEG, WN, and Blur.

TABLE XIV
PERFORMANCE OF THE TCLT-YCbCr METHOD WITH $K = 20$

Database	Metric	JP2K	JPEG	WN	Blur	FF	All
LIVE II	DIIVINE	0.913	0.910	**0.984**	0.921	0.863	0.916
	BLINDS-II	**0.951**	0.942	0.978	0.944	**0.927**	0.920
	BRIQUSE	0.914	**0.965**	0.979	0.951	0.877	0.940
	TCLT-YCbCr	0.917	0.925	0.981	**0.953**	0.921	**0.944**
TID2008	DIIVINE	0.924	0.866	**0.851**	0.862	--	0.889
	BLINDS-II	0.915	0.889	0.696	0.857	--	0.854
	BRIQUSE	0.832	**0.924**	0.829	**0.881**	--	**0.896**
	TCLT-YCbCr	**0.930**	0.878	0.839	0.864	--	0.879
CSIQ	DIIVINE	0.830	0.704	0.797	0.871	--	0.828
	BLINDS-II	0.884	0.881	0.886	0.870	--	0.873
	BRIQUSE	0.857	0.877	**0.926**	0.874	--	0.881
	TCLT-YCbCr	**0.900**	**0.915**	0.910	**0.898**	--	**0.892**

The detailed SROCC performances on TID2008 and CSIQ are presented in Tables XII and XIII, respectively. For comparison, we also show the performances of two FR metrics PSNR and SSIM in Tables XII and XIII. The performances of three BIQA metrics DIIVINE, BLINDS-II, and BRISQUE are also involved here, where the best BIQA metrics are highlighted in boldface. It is clear that all the four versions of the proposed method still achieve highly consistent evaluation with human perception. In addition, due to the contributions of multichannel feature fusion and distortion-specific LT, TCLT-YCbCr also performs best of all the four proposed methods for both TID2008 and CSIQ databases, which is consistent with the result in the LIVE II test.

Based on the experimental results here, we find that the four versions of the proposed method perform similarly on each database. If only a complete training set is properly constructed, the proposed four methods can accurately predict the perceptual quality of a degraded image.

F. Impact of the Neighbor Number

Here, we investigate the performance variation of the proposed method under different K, whose value ranges from 1 to 100 with an interval of 5. The investigation is implemented on the LIVE II database. Except for the K, all the other settings are the same as in Section IV-C. The median values of the RMSE across 100 train-test trials are reported in Fig. 11.

It is seen that when we change K from 1 to 20, the RMSE would significantly drop from 7.13 to 5.46. That is, a larger number of neighbors could avoid a large prediction error, which is caused by mistaking the neighbors' perceptual similarities. Then, the RMSE slowly increases and converges to 5.62 as K changes from 20 to 100. Here, too many neighbors would increase the model bias for introducing the dissimilar images. Since we employ the distance-based weight scheme, the predicted quality sores would not tend to the mean of the annotated database like the arithmetic averaging [64]. This can be found from Fig. 11, where our RMSE does not significantly increase as K increases to 100.

To evaluate the performance of our TCLT-YCbCr with 20 neighbors, we repeat the experiments in Sections IV-C and IV-E. The SROCC results on the LIVE II, TID2008, and CSIQ databases are reported in Table XIV, where the best results under each distortion type are highlighted in boldface. It is interesting to see that our method outperforms all the other BIQA metrics in the LIVE II and CSIQ databases. In the TID2008 database, our method still works well and achieves high consistency with human

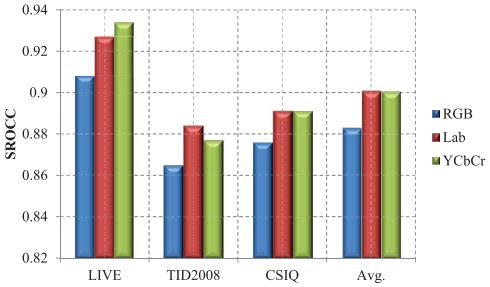


Fig. 12. Performance comparison for TCLT under different color spaces.

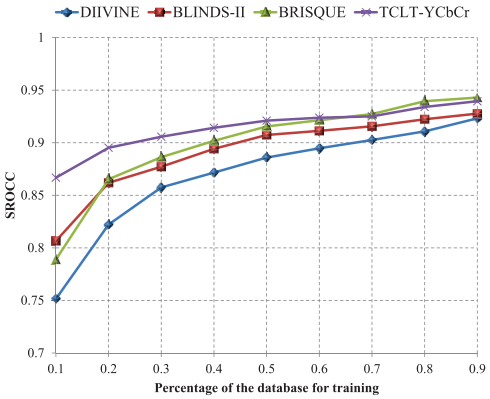


Fig. 13. Plot of median SROCC between the predicted quality index Q and DMOS versus the percentage of the images used for annotated set.

perception. Since the results in LIVE II are obtained from 100 train-test trials, we further perform a onesided *t-test* to evaluate the statistical significance of the difference between different BIQA methods. The reported result confirms that when K is set to 20, our TCLT-YCbCr is statistically superior to all of the DIIVINE, BLINDS-II, and BRISQUE metrics on the LIVE II database.

G. Comparison Among Different Color Space

Here, we further investigate the performance of our multichannel features under different color spaces, which include RGB, Lab, and YCbCr. We repeat the experiments in Sections IV-C and IV-E. For each candidate color space, the SROCC values on LIVE II, TID2008, and CSIQ and their average results across the three databases are reported in Fig. 12.

It is seen that the performance of RGB is inferior to those of Lab and YCbCr on the three databases. There are two possible reasons accounting for this result: first, both Lab and YCbCr represent the color image with one luma and two chroma components. Since the chroma components are insensitive to illumination change, it could improve the discriminative power of the features in describing the natural scene [65]. Second, by transforming from RGB to Lab/YCbCr, the high-frequency image content is preserved in the luma component and the smooth information is stored in the chroma components. Then, the statistics on Lab/YCbCr can capture more comprehensive intra-scale correlations from different color channels [66].

In addition, it can be seen that Lab and YCbCr achieve similar performances. As shown in Fig. 12, YCbCr slightly outperforms Lab in LIVE II. In TID, Lab performs better than YCbCr. For CSIQ, YCbCr achieves the same SROCC result as Lab. On average, the Lab and YCbCr also achieve the same performance across these three databases, whose average SROCCs are both 0.901.

TABLE XV
SROCC PERFORMANCE OF TCLT-YCbCr ON THE UNKNOWN SUBSETS OF THE TID2008 DATABASE

	ANC	SCN	MN	HFN	IN	QN	ID
BRISQUE	0.713	-0.495	0.583	0.623	0.582	0.739	0.570
TCLT-Gray	0.765	0.365	0.824	0.862	0.889	0.194	0.710
TCLT-Gray	0.733	0.490	0.816	0.863	0.799	-0.253	0.614
TCLT-YCbCr	0.818	0.625	0.830	0.882	0.827	0.681	0.673
TCLT-YCbCr	0.771	0.454	0.817	0.888	0.736	0.022	0.690
	JPEGTE	JP2KTE	NEPN	LBDDI	MS	CC	
BRISQUE	0.288	0.260	0.160	0.165	0.091	-0.052	
TCLT-Gray	0.018	-0.127	-0.154	0.014	0.110	0.008	
TCLT-Gray	-0.093	0.374	-0.153	0.139	0.233	-0.128	
TCLT-YCbCr	-0.180	-0.028	0.128	0.030	-0.069	0.095	
TCLT-YCbCr	-0.123	0.171	0.025	0.219	0.122	0.314	

H. Completeness of the Annotated Image Set

Here, we further discuss the completeness of the annotated samples. For our TCLT-YCbCr method, the completeness requires that the query image could find some similar images in the annotated sample set. It should satisfy two points: 1) the annotated set should cover all available DMOS values and 2) the annotated set should contain all possible distortion types.

1) Impact of the Annotated Sample Size: Here, we investigate the impact of the DMOS coverage by varying the annotated sample size. The experiment in Section IV-C is repeated with different train-test splits, where the percentage of train set ranges from 10% to median SROCC variation versus the percentage of train set is reported in Fig. 13, where the results of DIIVINE, BLINDS-II, and BRISQUE are also added for comparison.

It can be seen that our method outperforms all the other BIQA metrics under the train set size 10%–60%. For the train set size 70%–90%, our TCLT-YCbCr method is slightly inferior to the BRISQUE metric. When only 10% samples are used for training, our TCLT-YCbCr method can obtain a reasonable performance in the LIVE II database, whose median SROCC reaches up to 0.8668. In addition, we can also find that the SROCCs of all BIQA metrics monotonically go up as the train set size increases. Because a larger training set would cover more dense DMOS values, it makes the train set more complete and facilitates the LT.

2) Impact of the Unknown Distortion Type: Here, we implement the cross-database experiment to investigate the train set's completeness in terms of the distortion type. The LIVE II database is used as the annotated set. The test set is constructed with 13 unknown distortion types in TID2008, which are additive noise in color components (ANCs), spatially correlated noise (SCN), masked noise (MN), high-frequency noise (HFN), impulse noise (IN), quantization noise (QN), image denoising (ID), JPEG transmission errors (JPEGTE), JPEG2000 transmission errors (JP2KTE), noneccentricity pattern noise (NEPN), local block-wise distortions of different intensity (LBDDI), mean shift (MS), and contrast change (CC). The SROCCs of our methods are

TABLE XVI
COMPUTATIONAL COMPLEXITY FOR EACH FEATURE DOMAIN
(N: THE NUMBER OF PIXELS IN AN IMAGE)

Module	Percentage of Time (%)	Complexity	Note
DCT	78.43	$O\left(N/d^2 \log_2(N/d^2) + b \cdot N/d^2 + (2f-1)\cdot c\right)$	d: block size, b: no. of skewness bins f: no. of freq. bands, c: no. of coefficient bins
Wavelet	11.74	$O\left(N \log_2 N + (2\hat{f}-1)\cdot \hat{c}\right)$	$\hat{f}$: no. of subbands, $\hat{c}$: no. of coefficient bins
Spatial	7.73	$O(N)$	
LT	2.09	$O(N_r)$	N_r: no. of reference images

reported in Table XV. For comparison, BRISQUE [62] is also tested here.

Since the test images distortion types are not present in the annotated samples, it is hard for us to implement an efficient LT. For SCN, QN, JPEGTE, JP2KTE, NEPN, LBDDI, MS, and CC, their SROCCs are all smaller than 0.5 for TCLT-YCbCr. However, when the unknown distortion types show some characteristics similar to those of the annotated samples, the proposed method would still work well. For example, the ANC, HFN, IN, and MN are similar to the WN, and ID is similar to Blur. Correspondingly, our TCLT-YCbCr method achieves relatively high SROCC values on these distortion types, which are all larger than 0.65. Similar results can be found for BRISQUE. Experimental results show that the application of TCLT-YCbCr is limited to certain distortion types which are similar to the training data.

I. Computational Analysis

In this section, we further analyze the computational complexity of the proposed method. This investigation is conducted on the LIVE II database, whose 80% images are used for constructing the annotated reference samples. The system platform is an Intel Core 2 processor with a speed of 2.0 GHz, 2-GB RAM, and Windows 7. The unoptimized MATLAB code is run on the MATLAB R2009b software. Here, it takes about 10 s to predict the quality of a 512×768 image, which includes both the feature extraction and the LT steps.

To compare the computational loads of each module, both the percentage of running time and the time complexity for each module are summarized in Table XVI. Compared with the existing BIQA metrics [15], [16], the proposed method introduces an extra online prediction module LT. However, its complexity $O(N_r)$ is determined by the number of the annotated reference samples, which is negligible with respect to the total complexity. As shown in Table XVI, the LT module only takes 2% of running time in the proposed method.

In addition, the complexities listed in Table XVI show the upper bound of each module, which is computed in terms of the serial operation. In fact, the proposed algorithm is highly parallelizable, where the quality-aware information can be extracted from multiple domains simultaneously. Accordingly, the application of our proposed TCLT method should not suffer the limitation due to its complexity.

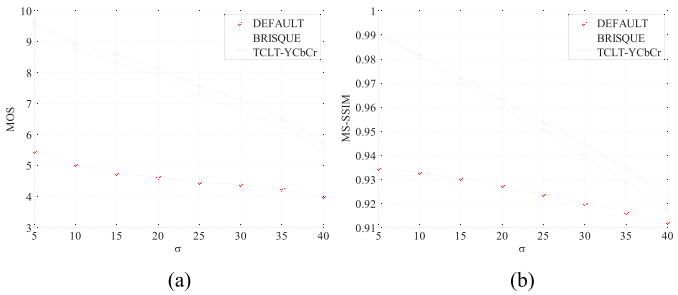


Fig. 14. Mean quality of the denoised images at each noise level. The x-axis denotes the noise variance σ and the y-axis denotes the quality metric. (a) Subjective metric MOS. (b) Objective metric MS-SSIM.

J. Discussion

Since the proposed DCT domain features are not shift invariant, a simple translation on the image may result in significant change on the features. An alternative solution is to add the shifted images in the annotated set. In our future work, more robust features will be studied. Similar to many previous works, we assume that a single distortion is present in the image, which is consistent with the condition in existing IQA databases. In many situations, an image may be contaminated by multiple sources, which brings greater difficulty for BIQA. Until now, the BIQA for the hybrid distortion is still a challenging problem and there are few reliable hybrid distortion IQA databases. In our future work, we would make efforts to build the hybrid distortion IQA data and extend the proposed method to deal with this new problem.

V. APPLICATION FOR IMAGE AUTO-DENOISING

To verify the efficiency of the proposed algorithm, we further investigate its performance in an image denoising application. Many representative denoising methods require some manual parameters (e.g., noise variance) to obtain good results. However, in practice, it is impractical to manually select the parameters for each test image. Here, the BIQA metric is quite desirable in estimating the perceptual quality of the denoised image, which could help us to automatically choose the denoising parameter.

We incorporate our TCLT-YCbCr metric into a representative color image denoising algorithm CBM3D [67], which needs the noise variance σ as the input parameter. In our implementation, the test image is first denoised with different parameters, and then the BIQA metric is used to evaluate the perceptual quality of each denoised image. By choosing the parameter that produces the highest image quality, we can obtain a good denoising result without user intervention. Here, the LIVE II database is still used as the annotated set for our TCLT-YCbCr method. To avoid overlapping between the annotated set and the test set, we build a denoised image dataset, whose 23 original images are extracted from the Computer Vision Group-University of Granada database (http://decsai.ugr.es/cvg/dbimagenes/). We add 8 levels of Gaussian noise to each original image, where σ can range from 5 to 40 at an interval of 5.

In the testing stage, the denoising results obtained by the default parameter [67] are used as the benchmark. For our

method, we run CBM3D with 50 candidate σ, which ranges from 1 to 50. Then, the denoised image which is perceived best for our proposed metric is selected as the output. For comparison, the denoising results chosen by the BRISQUE metric are also investigated here. In the subjective test, we invite 12 subjects to rate the perceptual quality for each denoised image, where the subjective score can range from 1 to 10 and a larger value denotes a better quality. Then, the MOS across 12 subjects is assigned to each denoised image. Besides the MOS, an objective metric MS-SSIM [68] is also used to measure the denoising results.

Here, we compute the mean quality of 23 denoised images at each distortion level to compare different methods. Both the subjective and objective evaluation results are shown in Fig. 14. It can be seen that the denoised image's quality could be significantly improved by selecting the parameter with the TCLT-YCbCr and BRISQUE metrics. In addition, in terms of both the MOS and MS-SSIM, the denoising results obtained from our TCLT-YCbCr are better than those of BRISQUE, as shown in Fig. 14. In addition, we also implement the one-sided *t-test* on the MOS and MS-SSIM of all denoised images. The reported result shows that our TCLT-YCbCr-based denoising results are statistically superior to those of BRISQUE and default methods in both the subjective and objective metrics.

VI. Conclusion

In this paper, we propose a novel blind image quality assessment algorithm based on distortion-TCLT. Inspired by the hierarchical and trichromatic properties of human vision, we extract both the frequency and spatial-frequency features from all three YCbCr channels to describe a natural image. Then, a KNN-based LT model is used to estimate the query image's quality. Both the validity and robustness of the proposed algorithm have been verified through the experiments on LIVE II, TID2008, and CSIQ databases. In addition, experimental results show that the application of the proposed method is limited to certain distortion types which are similar to the training data. It is still a challenging task to deal with the unknown distortion types.

In our future work, the mid-level features will be studied to simulate the more complex visual perception properties, e.g., shape selectivity and saliency. In addition, a more efficient similarity metric will be studied to further improve the performance of the LT module.

References

[1] Z. Wang, "Applications of objective image quality assessment methods [applications corner]," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 137–142, Nov. 2011.

[2] Z. Wang and A. C. Bovik, "Reduced- and no-reference image quality assessment," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 29–40, Nov. 2011.

[3] J. Jiang, K. Qiu, and G. Xiao, "A block-edge-pattern-based content descriptor in DCT domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 7, pp. 994–998, Jul. 2008.

[4] P.-C. Wu and L.-G. Chen, "An efficient architecture for two-dimensional discrete wavelet transform," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 4, pp. 536–545, Apr. 2001.

[5] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: Based on eye-tracking data," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 7, pp. 971–982, Jul. 2011.

[6] H. Li and K. N. Ngan, "A co-saliency model of image pairs," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3365–3375, Dec. 2011.

[7] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[8] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[9] Z. Gao and Y. F. Zheng, "Quality constrained compression using DWT-based image quality metric," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 7, pp. 910–922, Jul. 2008.

[10] J. Wu, W. Lin, G. Shi, and A. Liu, "Perceptual quality metric with internal generative mechanism," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 43–54, Jan. 2013.

[11] I. Gkioulekas, G. Evangelopoulos, and P. Maragos, "Spatial Bayesian surprise for image saliency and quality assessment," in *Proc. 17th IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 1081–1084.

[12] J. A. Redi, P. Gastaldo, I. Heynderickx, and R. Zunino, "Color distribution information for the reduced-reference assessment of perceived image quality," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1757–1769, Dec. 2010.

[13] R. Soundararajan and A. C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 517–526, Feb. 2012.

[14] M. A. Saad, A. C. Bovik, and C. Charrier, "A DCT statistics-based blind image quality index," *IEEE Signal Process. Lett.*, vol. 17, no. 6, pp. 583–586, Jun. 2010.

[15] M. A. 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, Aug. 2012.

[16] 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.

[17] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[18] H. Tang, N. Joshi, and A. Kapoor, "Learning a blind measure of perceptual image quality," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 305–312.

[19] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, May 2010.

[20] C. Li, A. C. Bovik, and X. Wu, "Blind image quality assessment using a general regression neural network," *IEEE Trans. Neural Netw.*, vol. 22, no. 5, pp. 793–799, May 2011.

[21] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Comput.*, vol. 12, no. 5, pp. 1207–1245, May 2000.

[22] D. F. Specht, "A general regression neural network," *IEEE Trans. Neural Netw.*, vol. 2, no. 6, pp. 568–576, Nov. 1991.

[23] S. Marcelja, "Mathematical description of the responses of simple cortical cells," *J. Opt. Soc. Amer.*, vol. 70, no. 11, pp. 1297–1300, Nov. 1980.

[24] D. M. MacKay, "Strife over visual cortical function," *Nature*, vol. 289, no. 5794, pp. 117–118, 1981.

[25] S. G. Solomon and P. Lennie, "The machinery of colour vision," *Nature Rev. Neurosci.*, vol. 8, no. 4, pp. 276–286, 2007.

[26] Q. Wu, H. Li, K. N. Ngan, B. Zeng, and M. Gabbouj, "No reference image quality metric via distortion identification and multi-channel label transfer," in *Proc. IEEE Int. Symp. Circuits Syst.*, Jun. 2014, pp. 530–533.

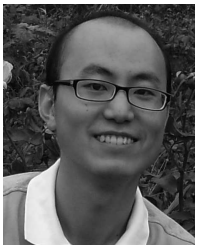
[27] N. Kruger *et al.*, "Deep hierarchies in the primate visual cortex: What can we learn for computer vision?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1847–1871, Aug. 2013.

[28] D. J. Felleman and D. C. Van Essen, "Distributed hierarchical processing in the primate cerebral cortex," *Cerebral Cortex*, vol. 1, no. 1, pp. 1–47, 1991.

[29] 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, accessed 2006.

[30] 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. Modern Radioelectron.*, vol. 10, no. 4, pp. 30–45, 2009.

[31] E. C. Larson and D. M. Chandler. *Categorical Image Quality (CSIQ) Database*. [Online]. Available: http://vision.okstate.edu/csiq, accessed 2010.
[32] D. Gao and N. Vasconcelos, "Discriminant saliency for visual recognition from cluttered scenes," in *Proc. Adv. Neural Inform. Process. Syst.*, 2004, pp. 481–488.
[33] J. Solomon, A. B. Watson, and A. Ahumada, "Visibility of DCT basis functions: Effects of contrast masking," in *Proc. Data Compress. Conf.*, Mar. 1994, pp. 361–370.
[34] H. B. Barlow, "Understanding natural vision," in *Physical and Biological Processing of Images*. Berlin, Germany: Springer-Verlag, 1983, pp. 2–14.
[35] C. Tu and T. D. Tran, "Context-based entropy coding of block transform coefficients for image compression," *IEEE Trans. Image Process.*, vol. 11, no. 11, pp. 1271–1283, Nov. 2002.
[36] P. T. von Hippel, "Mean, median, and skew: Correcting a textbook rule," *J. Statist. Edu.*, vol. 13, no. 2, Jul. 2005.
[37] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer. A*, vol. 4, no. 12, pp. 2379–2394, 1987.
[38] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.
[39] M. Kristan, J. Perš, M. Perše, and S. Kovačič, "A Bayes-spectral-entropy-based measure of camera focus using a discrete cosine transform," *Pattern Recognit. Lett.*, vol. 27, no. 13, pp. 1431–1439, 2006.
[40] Z. He, W. Lu, W. Sun, and J. Huang, "Digital image splicing detection based on Markov features in DCT and DWT domain," *Pattern Recognit.*, vol. 45, no. 12, pp. 4292–4299, 2012.
[41] B. D. B. Willmore, R. J. Prenger, and J. L. Gallant, "Neural representation of natural images in visual area V2," *J. Neurosci.*, vol. 30, no. 6, pp. 2102–2114, 2010.
[42] Y. Karklin and M. S. Lewicki, "Emergence of complex cell properties by learning to generalize in natural scenes," *Nature*, vol. 457, no. 7225, pp. 83–86, 2009.
[43] M. Gavish, B. Nadler, and R. R. Coifman, "Multiscale wavelets on trees, graphs and high dimensional data: Theory and applications to semi supervised learning," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 367–374.
[44] J. Liu and P. Moulin, "Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients," *IEEE Trans. Image Process.*, vol. 10, no. 11, pp. 1647–1658, Nov. 2001.
[45] B.-J. Kim and W. A. Pearlman, "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)," in *Proc. Data Compress. Conf.*, May 1997, pp. 251–260.
[46] Y. El-Shamayleh and J. A. Movshon, "Neuronal responses to texture-defined form in macaque visual area V2," *J. Neurosci.*, vol. 31, no. 23, pp. 8543–8555, 2011.
[47] L. G. Appelbaum, A. R. Wade, V. Y. Vildavski, M. W. Pettet, and A. M. Norcia, "Cue-invariant networks for figure and background processing in human visual cortex," *J. Neurosci.*, vol. 26, no. 45, pp. 11695–11708, 2006.
[48] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
[49] A. Ford and A. Roberts, *Colour Space Conversions*, vol. 1998. London, U.K.: Westminster Univ., 1998, pp. 1–31.
[50] J. Huang and Y. Wang, "Compression of color facial images using feature correction two-stage vector quantization," *IEEE Trans. Image Process.*, vol. 8, no. 1, pp. 102–109, Jan. 1999.
[51] S. Fan, T.-T. Ng, J. S. Herberg, B. L. Koenig, C. Y.-C. Tan, and R. Wang, "An automated estimator of image visual realism based on human cognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2014, pp. 4201–4208.
[52] Y. Agam and R. Sekuler, "Interactions between working memory and visual perception: An ERP/EEG study," *Neuroimage*, vol. 36, no. 3, pp. 933–942, 2007.
[53] R. L. Buckner, M. E. Raichle, F. M. Miezin, and S. E. Petersen, "Functional anatomic studies of memory retrieval for auditory words and visual pictures," *J. Neurosci.*, vol. 16, no. 19, pp. 6219–6235, 1996.
[54] G. A. F. Seber and C. J. Wild, *Nonlinear Regression* (Probability and Statistics). New York, NY, USA: Wiley, 2003.
[55] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, May 2011, Art. ID 27.
[56] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," Dept. Comput. Sci., Nat. Taiwan Univ., Taipei, Taiwan, Tech. Rep. V-901–V-904, 2003. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf
[57] P. Zhang, "Model selection via multifold cross validation," *Ann. Statist.*, vol. 21, no. 1, pp. 299–313, 1993.
[58] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.
[59] A. Kale, A. K. Roychowdhury, and R. Chellappa, "Fusion of gait and face for human identification," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5. May 2004, pp. V-901–V-904.
[60] A. Mittal, G. S. Muralidhar, J. Ghosh, and A. C. Bovik, "Blind image quality assessment without human training using latent quality factors," *IEEE Signal Process. Lett.*, vol. 19, no. 2, pp. 75–78, Feb. 2012.
[61] X. Gao, F. Gao, D. Tao, and X. Li, "Universal blind image quality assessment metrics via natural scene statistics and multiple kernel learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 12, pp. 2013–2026, Dec. 2013.
[62] 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.
[63] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*. Boca Raton, FL, USA: CRC Press, 2003.
[64] S. A. Dudani, "The distance-weighted k-nearest-neighbor rule," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, no. 4, pp. 325–327, Apr. 1976.
[65] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1582–1596, Sep. 2010.
[66] N.-X. Lian, V. Zagorodnov, and Y.-P. Tan, "Color image denoising using wavelets and minimum cut analysis," *IEEE Signal Process. Lett.*, vol. 12, no. 11, pp. 741–744, Nov. 2005.
[67] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Color image denoising via sparse 3D collaborative filtering with grouping constraint in luminance-chrominance space," in *Proc. IEEE Int. Conf. Image Process.*, vol. 1. Sep./Oct. 2007, pp. I-313–I-316.
[68] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th IEEE Asilomar Conf. Signals, Syst. Comput.*, vol. 2. Nov. 2003, pp. 1398–1402.



Qingbo Wu (S'12–M'13) received the B.Eng. degree in education of applied electronic technology from Hebei Normal University, Hebei, China, in 2009. He is currently working toward the Ph.D. degree with the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu, China.

He joined the Image and Video Processing Laboratory, The Chinese University of Hong Kong, Hong Kong, in 2014, as a Research Assistant. He is currently a Visiting Student with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research interests include image/video coding, quality evaluation, and perceptual modeling and processing.

Hongliang Li (SM'12) received the Ph.D. degree in electronics and information engineering from Xi'an Jiaotong University, Xi'an, China, in 2005.

He joined the Visual Signal Processing and Communication Laboratory, The Chinese University of Hong Kong, Hong Kong, from 2005 to 2006, as a Research Associate, where he was a Post-Doctoral Fellow from 2006 to 2008. He was involved in many professional activities. He is currently a Professor with the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu, China. He has authored or co-authored numerous technical articles in well-known international journals and conferences. His research interests include image segmentation, object detection, image and video coding, visual attention, and multimedia communication system.

Dr. Li is a Co-Editor of a book entitled *Video Segmentation and Its Applications* (Springer). He is a member of the Editorial Board of *Journal on Visual Communications and Image Representation* and is the Area Editor of *Signal Processing: Image Communication* (Elsevier Science). He served as a Technical Program Co-Chair of the International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS) in 2009, the General Co-Chair of ISPACS in 2010, the Publicity Co-Chair of the IEEE Visual Communications and Image Processing (VCIP) in 2013, the Local Chair of the IEEE International Conference on Multimedia and Expo (ICME) in 2014, and a TPC Member in a number of international conferences, e.g., ICME in 2012 and 2013, the IEEE International Symposium on Circuits and Systems in 2013, the Pacific-Rim Conference on Multimedia in 2007 and 2009, and VCIP in 2010. He will serve as a Technical Program Co-Chair of the IEEE VCIP 2016. He was selected as the New Century Excellent Talents in University by the Chinese Ministry of Education, China, in 2008.

Fanman Meng (S'12–M'14) received the Ph.D. degree in signal and information processing from University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2014.

He was with the Division of Visual and Interactive Computing, Nanyang Technological University, Singapore, from 2013 to 2014, as a Research Assistant. He is currently an Associate Professor with the School of Electronic Engineering, UESTC. He has authored or co-authored numerous technical articles in well-known international journals and conferences. His research interests include image segmentation and object detection.

Dr. Meng is a member of the IEEE Circuits and Systems Society. He received the Best Student Paper Honorable Mention Award for the 12th Asian Conference on Computer Vision, Singapore, in 2014, and the Top 10% Paper Award in the IEEE International Conference on Image Processing, Paris, France, in 2014.

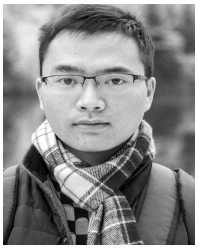
King N. Ngan (F'00) received the Ph.D. degree in electrical engineering from Loughborough University, Loughborough, U.K.

He was a Full Professor with Nanyang Technological University, Singapore, and University of Western Australia, Crawley WA, Australia. He has been the Chair Professor with University of Electronic Science and Technology, Chengdu, China, under the National Thousand Talents Program, since 2012. He is currently the Chair Professor with the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong. He holds honorary and visiting professorships in numerous universities in China, Australia, and Southeast Asia. He has published extensively, including three authored books, seven edited volumes, over 350 refereed technical papers, and edited nine special issues in journals. He holds 15 patents in image/video coding and communications.

Prof. Ngan is a fellow of the Institution of Engineering and Technology in U.K., and the Institute of Engineers Australia in Australia. He was the IEEE Distinguished Lecturer from 2006 to 2007. He served as an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, *Journal of Visual Communications and Image Representation*, *EURASIP Journal of Signal Processing: Image Communication*, and *Journal of Applied Signal Processing*. He was the Chair and Co-Chair of a number of prestigious international conferences on image and video processing, including the 2010 IEEE International Conference on Image Processing, and served on the Advisory and Technical Committees of numerous professional organizations.

Bing Luo received the B.Sc. degree in communication engineering from The Second Artillery Command College, in 2009, and the M.Sc. degree in computer application technology from Xihua University, in 2012, respectively. He is currently pursuing the Ph.D. degree with the Intelligent Visual Information Processing and Communication Laboratory, University of Electronic Science and Technology of China, Chengdu, China, under the supervision of Prof. H. Li.

His current research interests include image and video segmentation, and machine learning.



Chao Huang received the B.Sc. degree from University of Electronic Science and Technology of China, Chengdu, China, in 2012, under the supervision of Prof. H. Li. He is currently working toward the Ph.D. degree with the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu, China.

His research interests include image classification and segmentation.

Bing Zeng (M'91–SM'13) received the B.Eng. and M.Eng. degrees from University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1983 and 1986, respectively, and the Ph.D. degree from Tampere University of Technology, Tampere, Finland, in 1991, all in electrical engineering.

He was a Post-Doctoral Fellow with University of Toronto, Toronto, ON, Canada, from 1991 to 1992 and was a Researcher with Concordia University, Montréal, QC, Canada, from 1992 to 1993. He was a Visiting Researcher with Microsoft Research Asia, Beijing, China, in 2000. He joined the Hong Kong University of Science and Technology, Hong Kong, in 1993, where he is currently a Full Professor with the Department of Electronic and Computer Engineering. Since 2013, he has been a National 1000-Talent-Program Chair Professor with UESTC, where he leads the Institute of Image Processing to work on image/video coding and processing, 3-D/multiview video technology, and visual big data. His research activities have generated over 200 publications in various leading journals and conferences. His current research interests include image/video coding and processing, image superresolution, endoscopy image/video processing, compressive sensing theory and applications, light-field signal processing, and HDTV technology.

Prof. Zeng's three representing works are as follows. One paper on fast block motion estimation, published in IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY in 1994, has been SCI-cited more than 870 times (Google-cited more than 1800 times), and currently stands at the seventh position among all the papers published in this IEEE journal (since its launch in 1991). Another paper on smart padding for arbitrarily shaped image blocks, published in IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY in 2001, has led to a U.S. patent that has been successfully licensed to the industry. His third paper on directional transforms, published in IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY in 2008, received the 2011 IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY Best Paper Award. He was a recipient of the Second Class Natural Science Award (the first recipient) from the Chinese Ministry of Education in 2014, and the best paper award at ChinaCom three times (2009 Xi'an, 2010 Beijing, and 2012 Kunming). He served as an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from 1995 to 1999 and 2010 to 2014. He is on the Editorial Board of *Journal of Visual Communication and Image Representation*. He served in various positions in a number of international conferences. He is a member of the Visual Signal Processing and Communications Technical Committee of the IEEE Circuits and Systems Society and the Multimedia Communications Technical Committee of the IEEE Communications Society. He will be the General Chair of the IEEE Visual Communications and Image Processing, Chengdu, in 2016.