

# Perceptual Hashing With Complementary Color Wavelet Transform and Compressed Sensing for Reduced-Reference Image Quality Assessment

Mengzhu Yu<sup>ID</sup>, Zhenjun Tang<sup>ID</sup>, *Member, IEEE*, Xianquan Zhang<sup>ID</sup>, Bineng Zhong<sup>ID</sup>,  
and Xinpeng Zhang<sup>ID</sup>, *Member, IEEE*

**Abstract**—Image quality assessment (IQA) is an important task of image processing and has diverse applications, such as image super-resolution reconstruction, image transmission and monitoring systems. This paper proposes a perceptual hashing algorithm with complementary color wavelet transform (CCWT) and compressed sensing (CS) for reduced-reference (RR) IQA. The CCWT is exploited to decompose input color image into different sub-bands. Since the calculation of CCWT uses all color channels without discarding any information, the distortions introduced by digital operations on color channels are preserved in the CCWT sub-bands. The block-based CS is used to extract features from the CCWT sub-bands. As the Euclidean distance between the block-based CS features is slightly influenced by content-preserving operations, perceptual features constructed by Euclidean distances are robust, discriminative and compact. Hash sequence is finally determined by quantifying the perceptual features. Effectiveness of the proposed hashing is verified by various experiments on four open image databases. Experimental results demonstrate that the proposed hashing is superior to some state-of-the-art algorithms in terms of classification and RR IQA application.

**Index Terms**—Image quality assessment, image hashing, complementary color wavelet transform (CCWT), compressed sensing (CS), Sobel operator.

## I. INTRODUCTION

WITH the advent of the big data era, the number of images is rapidly increasing and the demands for image

Manuscript received 28 April 2022; revised 21 June 2022; accepted 4 July 2022. Date of publication 13 July 2022; date of current version 28 October 2022. This work was supported in part by the Guangxi Natural Science Foundation under Grant 2022GXNSFAA035506; in part by the Science and Technology Project of Guangxi under Grant GuiKeAD21220114; in part by the National Natural Science Foundation of China under Grant 61962008, Grant 62062013, and Grant U1936214; in part by the Guangxi “Bagui Scholar” Team for Innovation and Research; in part by the Guangxi Talent Highland Project of Big Data Intelligence and Application; and in part by the Innovation Project of Guangxi Graduate Education under Grant YCBZ2022063. This article was recommended by Associate Editor B. Wen. (Corresponding author: Zhenjun Tang.)

Mengzhu Yu, Zhenjun Tang, Xianquan Zhang, and Bineng Zhong are with the Guangxi Key Laboratory of Multi-Source Information Mining and Security, Guangxi Collaborative Innovation Center of Multi-Source Information Integration and Intelligent Processing, Guangxi Normal University, Guilin 541004, China (e-mail: yumz1030@163.com; tangzj230@163.com; zqx6622@163.com; bnzhong@gxnu.edu.cn).

Xinpeng Zhang is with the School of Computer Science, Fudan University, Shanghai 200433, China (e-mail: zhangxinpeng@fudan.edu.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TCSVT.2022.3190273>.

Digital Object Identifier 10.1109/TCSVT.2022.3190273

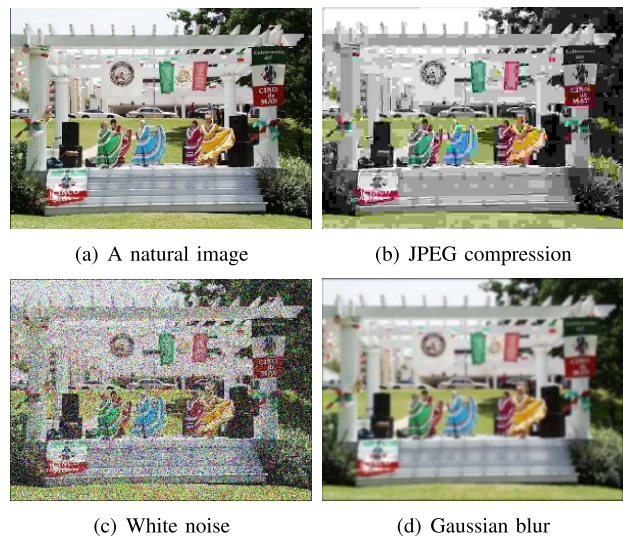


Fig. 1. A natural image and its three distorted versions.

quality assessment (IQA) have soared in many applications, such as image super-resolution reconstruction, image transmission and monitoring systems. For example, many image acquisition systems require a useful IQA scheme to adjust system parameters for obtaining good image quality. Therefore, it is highly desired to grade image quality in real-time for maintaining the required quality of service. Consequently, it is an important task to develop efficient IQA schemes for diverse image applications [1]–[3]. Figure 1 illustrates a practical example of three types of image distortion, where (a) a natural image, (b) is a compressed version after JPEG compression, (c) is a noise version attacked by white noise, and (d) is a blurred version after Gaussian blurring. Efficient IQA schemes are expected to perceive distortion levels of these images. In general, the existing schemes of objective IQA can be classified into three kinds as follows. (i) Full-reference (FR) IQA: The inputs are the reference image and its distorted image. (ii) Reduced-reference (RR) IQA: The inputs are representative information of the reference image and its distorted image. (iii) No-reference (NR) IQA: The input is only the distorted image. In recent years, some researchers have proposed to solve RR IQA problem via perceptual image

hashing [4], [5]. This paper investigates a novel perceptual hashing algorithm for RR IQA.

Image hashing [6], [7] is a useful technology of image processing. It can effectively extract a visual content-based compact hash sequence from input image, which can be used to represent the input image itself. Since compact hash sequence has the advantages of low storage and fast calculation of similarity, image hashing has been widely applied to image copy detection, image retrieval, image tampering detection and image content authentication [8]–[10]. Robustness and discrimination are two major properties of image hashing. Specifically, robustness requires that the hash sequence of the original image and that of its distorted version are similar. Discrimination requires that hash sequences of different images are entirely distinct. The two properties are interrelated and mutually constrained. An effective algorithm of image hashing should make a desirable balance between them. In addition, for the application of RR IQA, perceptual image hashing must satisfy distortion sensitivity, which means that hashing algorithm needs to perceive the level of distortion on similar images. Note that there is a mutual constraint between robustness and distortion sensitivity. Strong robustness will lead to insensitivity of measuring distortion of similar image, while high distortion sensitivity will decrease the robustness. Therefore, it is a challenging task to develop novel perceptual hashing algorithms with a good trade-off between robustness and distortion sensitivity for the application of RR IQA.

Currently, most image hashing algorithms only consider sufficient robustness. They often ignore distortion sensitivity, and thus do not reach a desirable performance in the application of RR IQA. To tackle this problem, the paper proposes a novel perceptual hashing algorithm with complementary color wavelet transform (CCWT) and compressed sensing (CS). Compared with the existing hashing algorithms, there are three contributions as follows:

(1) The CCWT is exploited to decompose input color image into different sub-bands for extracting perceptual features. Since the calculation of CCWT uses all color channels without discarding any information, the distortions introduced by digital operations on color channels are preserved in the CCWT sub-bands. Since the low-frequency sub-bands contain basic image information, the features extracted from these sub-bands can distinguish images with different contents. As the high-frequency sub-bands reflect multi-directional color information, the color features extracted from these sub-bands can describe image distortion on different color channels.

(2) Perceptual features are extracted from the CCWT sub-bands via block-based CS. Since CS can directly achieve compression coding during the sampling process, the block-based CS features retain original feature distortion and provide good distortion sensitivity of our hashing algorithm. As the Euclidean distance between the block-based CS features is slightly influenced by content-preserving operations, perceptual features constructed by Euclidean distances are robust, discriminative and compact.

(3) Numerous experiments with 140990 images (100350 for robustness analysis and 40640 for discrimination evaluation) are done to evaluate the performances of the proposed hashing

algorithm. The results demonstrate that our hashing algorithm not only has good robustness and strong discrimination, but also keeps a good balance between them. In addition, the LIVE and TID2013 databases are used to validate our application in RR IQA. The experimental results show that the proposed algorithm outperforms some well-known IQA schemes.

The rest of this paper is organized as follows. Section II gives a review of the related work. Section III provides a specific description of the proposed hashing algorithm. Section IV describes the experimental results and performance comparisons. Our application in RR IQA is demonstrated in Section V. Finally, Section VI gives the conclusions.

## II. RELATED WORK

Various image hashing algorithms have been developed by researchers. According to their feature extraction techniques, the existing hashing algorithms can be categorized into the following four types.

### A. Transform Domain Based Hashing Algorithms

Robust features of these hashing algorithms are extracted from transform domain via transform techniques. The discrete wavelet transform (DWT), discrete cosine transform (DCT), Fourier-Mellin transform (FMT), Log-Polar transform (LPT) and Radon transform (RT) are popular transform techniques. For example, Venkatesan *et al.* [11] first utilized a wavelet representation of DWT to derive hash. This algorithm is robust against normal operations, but it is insensitive to malicious tampering. Wang *et al.* [12] jointly used DWT and DCT to extract features and compressed features by Karhunen–Loeve transformation. This algorithm is robust to normal digital operations, except rotation. Ahmed *et al.* [13] used DWT coefficients to generate the intermediate hash and compressed it via the SHA-1 function. This algorithm can find local tampering areas, but it is fragile to contrast and brightness adjustment. Qin *et al.* [14] used non-uniform sampling to extract frequency features in discrete Fourier domain. This algorithm only make use of the luminance component of images. Swaminathan *et al.* [15] encrypted features by a key and thus designed the FMT-based hashing algorithm for improving hash security. In [16], image features with rotation invariance were extracted by LPT. The two algorithms [15], [16] are both robust against rotation within  $10^\circ$ . Liu *et al.* [17] exploited RT and invariant features to design a hashing algorithm. Since this algorithm contains multiple transformations, its computational complexity is not satisfactory.

### B. Visual Saliency Detection Based Hashing Algorithms

Many visual saliency detection methods have been incorporated into the image hashing research for improving the performance of robustness. For instance, Monga *et al.* [18] developed a hashing algorithm with visual salient feature points. This algorithm can resist JPEG compression, but it is not resilient to geometric distortions. Wang *et al.* [19] extracted Gabor features for constructing visual system-based

hashing. This hashing has a good classification performance, but it is insensitive to malicious changes in small blocks. In [20], a global Zernike moment and visual attention detection were utilized to design a hashing algorithm for content authentication. Wang *et al.* [21] detected visually important features by Watson’s attention model and combined features with key points for image content authentication. This algorithm has good robustness, but it does not consider contrast adjustment and watermark embedding. In [22], the Phase spectrum of Fourier Transform (PFT) model of saliency detection and the ring partition were both utilized to develop a hashing algorithm. This hashing is resilient against large-angle rotation.

### C. Matrix Factorization Based Hashing Algorithms

Some commonly-used dimension reduction techniques are introduced to the research of image hashing. For example, Kozat *et al.* [23] employed singular value decomposition (SVD) to derive hash for ensuring robustness. But the robustness improvement tends to increase misclassification. Motivated by the SVD-based hashing [23], non-negative matrix factorization (NMF) was applied to the design of image hashing [24]. This hashing algorithm can resist JPEG compression and rotation. But its robustness against watermark embedding must be increased. The hashing algorithm with CS and visual information fidelity was designed by Kang *et al.* [25]. This algorithm demonstrates good robustness against some operations. However, its discrimination performance is not good. In [4], Lv and Wang extracted features via fast Johnson-Lindenstrauss transform to generate hash for RR IQA. This algorithm only considers the compression distortions of JPEG and JPEG2000. In another work, tensor decomposition is first used for designing image hashing by Tang *et al.* [8]. Their hashing uses image blocks to construct a tensor and exploits Tucker decomposition to generate hash. However, the hashing only resists rotation with a small angle.

### D. Statistical Features Based Hashing Algorithms

Different statistical approaches are adopted to extract features in image hashing. For example, Tang *et al.* [26] selected statistical features of an image ring as a feature vector and calculated invariant distance between feature vectors for hash generation. The ring-based algorithm can be robust to any-angle rotation. Qin *et al.* [27] used three techniques to extract features with structural information, including SVD, Canny operator and color vector angle. This algorithm is resistant to filtering and compression. Tang *et al.* [5] combined Canny operator with Weighted DWT (WDWT) statistical features for hash construction. This hashing provides a good measure of perceptual image distortion. But the dataset used for validating IQA performance is small. Huang *et al.* [1] developed a hashing algorithm for RR IQA of screen content images. This algorithm utilized local features of the gradient magnitude map and the normalized histogram for hash construction. Recently, Singh *et al.* [28] combined KAZE features and statistical features from the reference image to derive hash. This algorithm reaches good robustness and discriminative capabilities.

TABLE I  
SUMMARY OF NOTATIONS

$U$ : image size	$\mathbf{D}_1$ : low-frequency feature sequence
$U_l$ : sub-band size	$b$ : block size for low-frequency
$\mathbf{P}^{(L)}$ : low-frequency sub-band	$F$ : block number for low-frequency
$\mathbf{P}_{\text{total}}^{(R-C)}$ : Red-Cyan sub-band	$\mathbf{P}_{\text{total}}^{(W-B)}$ : White-Black sub-band
$\mathbf{P}_{\text{total}}^{(B-Y)}$ : Blue-Yellow sub-band	$\mathbf{P}_{\text{total}}^{(G-M)}$ : Green-Magenta sub-band
$\mathbf{D}_2$ : color feature sequence	$k$ : block size for color feature
$\mathbf{Y}$ : measurement matrix	$E$ : block number for color feature
$\mathbf{h}$ : hash sequence	$\mathbf{G}$ : maximum gradient map

In addition to the algorithms mentioned above, other techniques are also applied to image hashing. For instance, random Gabor filtering (GF) and dithered lattice vector quantization (LVQ) were jointly exploited to develop an algorithm [29]. This hashing algorithm is robust against JPEG compression and rotation, but its discrimination is unsatisfactory. The random walk (RW) hashing [30] was designed by adopting zigzag blocks with RW. This hashing enhances security via RW, but it is sensitive to rotation and translation. In [31], sparse coding was used to find features for hash construction. A fast discrete collaborative multimodal hashing algorithm was developed for image retrieval [32]. This hashing preserves high-level semantics while keeping low-level data features.

The above review shows that the current hashing algorithms have made a considerable progress. However, most hashing algorithms only consider sufficient robustness. They often ignore distortion sensitivity, and thus do not reach a desirable performance in the application of RR IQA. To tackle this problem, we design a perceptual image hashing algorithm which can reach good classification performance and is effective in RR IQA.

## III. PROPOSED HASHING ALGORITHM

The block diagram of the proposed hashing algorithm is depicted in Figure 2. Our hashing algorithm is composed of the following six steps: (i) Pre-processing operations are applied to input image, where bi-cubic interpolation is exploited to convert the size of input image to  $U \times U$ , and Gaussian low-pass filtering is used to eliminate noise. (ii) Two-level 2D CCWT is applied to the preprocessed image, and a low-frequency sub-band and four complementary color sub-bands are then generated for feature extraction. (iii) Low-frequency features are extracted from the low-frequency sub-band via block-based CS. (iv) Multi-directional edge detection is utilized to extract image edges from complementary color sub-bands. (v) Color features are extracted from maximum edge gradient maps via block-based CS. (vi) The hash sequence is constructed by concatenating the quantized low-frequency features and color features. Details of our hashing algorithm are described in the below subsections. Besides, Table I summarizes the relevant notations used in the paper.

### A. 2D CCWT

The classical wavelet transform has been widely applied to many applications, such as IQA [5], image fusion [33]

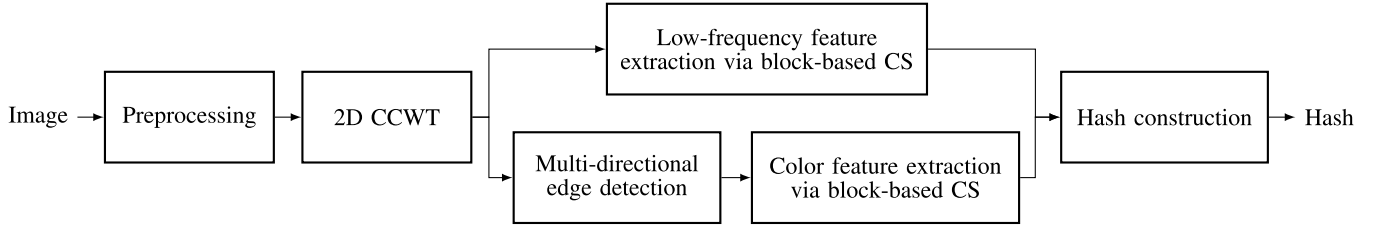


Fig. 2. Block diagram of the proposed hashing algorithm.

and image watermarking [34]. To improve the effectiveness of color image processing with wavelet transform, Chen *et al.* [35] firstly introduced color relations based on complementary visual theory into wavelet transform, and thus developed an innovative tool called CCWT. When the CCWT is used to process color image, it can focus on the advantages of traditional wavelets without discarding color-related information.

Complementary colors play key roles in human vision. Generally, they are represented as a pair of colors whose mixture is a white color. For example, white and black, red and cyan, green and magenta, and blue and yellow are four important pairs of complementary colors. It is known that complementary colors can be represented in the RGB hue ring [35], where the R, G and B axes are positioned at angles of  $0$ ,  $\frac{2\pi}{3}$  and  $\frac{4\pi}{3}$  along the hue ring, respectively. In order to extend R, G and B axes to wavelet domain, the 1D CCWT with relative phase differences of  $\frac{2\pi}{3}$  is designed, i.e.,  $\phi^{(0)}$ ,  $\phi^{(2\pi/3)}$  and  $\phi^{(4\pi/3)}$ . Then, referring to classical DWT, the 2D single-channel CCWT is extended from the 1D CCWT. Specifically, 2D single-channel CCWT can be seen as the calculations of 1D CCWT along with horizontal and vertical directions. Since 1D CCWT has three different low-pass and/or high-pass filters, 2D single-channel CCWT is achieved by 9 DWT decompositions with varying phase combinations.

After decomposition, high-frequency components are available and then two-dimensional components are further created by calculating permutation and combination. Then the non-zero components are filtered out to obtain the two-dimensional wavelet groups, which have approximately eight directions  $n = \frac{m\pi}{8}$  ( $m = 1, 2, \dots, 8$ ) and three relative phase differences, i.e.,  $0$ ,  $\frac{2\pi}{3}$  and  $\frac{4\pi}{3}$ . Therefore, the 2D single-channel CCWT inherits the directional filtering characteristics from the traditional wavelet and has richer directional selectivity. More details of the 2D single-channel CCWT can be referred to [35].

When the 2D single-channel CCWT is applied to color images, the R, G and B channels are mapped to the corresponding  $0$ ,  $\frac{2\pi}{3}$  and  $\frac{4\pi}{3}$  wavelet bases, respectively. To do this, the relative phase difference wavelets  $\phi_{l,n}^{(0)}$ ,  $\phi_{l,n}^{(2\pi/3)}$  and  $\phi_{l,n}^{(4\pi/3)}$  of the CCWT with level  $l$  and direction  $n$  are represented as  $\phi_{l,n}^{(R)}$ ,  $\phi_{l,n}^{(G)}$  and  $\phi_{l,n}^{(B)}$ . Next, wavelet decomposition of RGB channels is performed by them, and the obtained wavelet coefficient vectors can be expressed as  $\mathbf{c}_{l,n}^{(R)} = \mathbf{r} * \phi_{l,n}^{(R)}$ ,  $\mathbf{c}_{l,n}^{(G)} = \mathbf{g} * \phi_{l,n}^{(G)}$  and  $\mathbf{c}_{l,n}^{(B)} = \mathbf{b} * \phi_{l,n}^{(B)}$ , respectively, in which  $\mathbf{r}$ ,  $\mathbf{g}$  and  $\mathbf{b}$  represent the R, G and B channel vectors and  $*$  is the

convolution operation. The complementary color operators are determined by the equations (1)-(4) as follows:

1) *White-Black Operator*:

$$\mathbf{P}_{l,n}^{(W-B)} = \mathbf{c}_{l,n}^{(R)} + \mathbf{c}_{l,n}^{(G)} + \mathbf{c}_{l,n}^{(B)} \quad (1)$$

2) *Red-Cyan Operator*:

$$\mathbf{P}_{l,n}^{(R-C)} = \mathbf{c}_{l,n}^{(R)} - \mathbf{c}_{l,n}^{(G)} - \mathbf{c}_{l,n}^{(B)} \quad (2)$$

3) *Green-Magenta Operator*:

$$\mathbf{P}_{l,n}^{(G-M)} = \mathbf{c}_{l,n}^{(G)} - \mathbf{c}_{l,n}^{(R)} - \mathbf{c}_{l,n}^{(B)} \quad (3)$$

4) *Blue-Yellow Operator*:

$$\mathbf{P}_{l,n}^{(B-Y)} = \mathbf{c}_{l,n}^{(B)} - \mathbf{c}_{l,n}^{(R)} - \mathbf{c}_{l,n}^{(G)} \quad (4)$$

Note that the  $\mathbf{P}_{l,n}^{(W-B)}$  can be considered as the chroma operator and it will be changed correspondingly when the chroma value changes. The  $\mathbf{P}_{l,n}^{(R-C)}$ ,  $\mathbf{P}_{l,n}^{(G-M)}$  and  $\mathbf{P}_{l,n}^{(B-Y)}$  reflect the changes in three complementary colors. These four operators can be represented uniformly in the matrix form as shown in the equation (5).

$$\begin{bmatrix} \mathbf{P}_{l,n}^{(W-B)} \\ \mathbf{P}_{l,n}^{(R-C)} \\ \mathbf{P}_{l,n}^{(G-M)} \\ \mathbf{P}_{l,n}^{(B-Y)} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{c}_{l,n}^{(R)} \\ \mathbf{c}_{l,n}^{(G)} \\ \mathbf{c}_{l,n}^{(B)} \end{bmatrix} \quad (5)$$

In this paper, two-level 2D CCWT is applied to the pre-processed image. Here, we take the mean of low-frequency components as the low-frequency sub-band  $\mathbf{P}^{(L)}$  and the sum of complementary color operators with  $l = 2$  as the final complementary color sub-bands, i.e., White-Black complementary color sub-band  $\mathbf{P}_{\text{total}}^{(W-B)} = \sum_{m=1}^8 \left| \mathbf{P}_{2,m\pi/8}^{(W-B)} \right|$ , Red-Cyan complementary color sub-band  $\mathbf{P}_{\text{total}}^{(R-C)} = \sum_{m=1}^8 \left| \mathbf{P}_{2,m\pi/8}^{(R-C)} \right|$ , Green-Magenta complementary color sub-band  $\mathbf{P}_{\text{total}}^{(G-M)} = \sum_{m=1}^8 \left| \mathbf{P}_{2,m\pi/8}^{(G-M)} \right|$  and Blue-Yellow complementary color sub-band  $\mathbf{P}_{\text{total}}^{(B-Y)} = \sum_{m=1}^8 \left| \mathbf{P}_{2,m\pi/8}^{(B-Y)} \right|$ . The sizes of these sub-bands are all  $U_l \times U_l$ , in which  $U_l = U/2^l$ . The CCWT coefficients in the low-frequency sub-band contain approximation information of image, while the CCWT coefficients in complementary color sub-bands contain multi-direction color information of image. Therefore, image features extracted from these sub-bands can not only distinguish different images but also indicate image distortion. A visual example of two-level 2D CCWT is presented in Figure 3, where (a) is a color

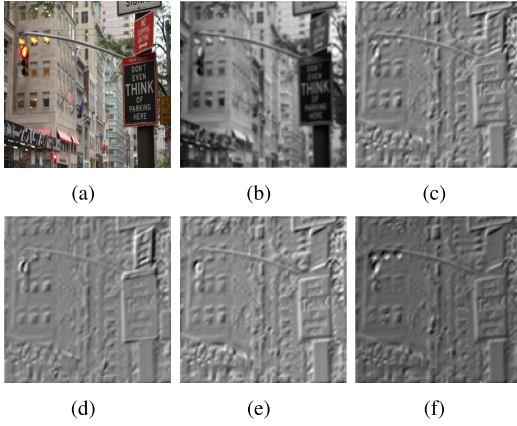


Fig. 3. A visual example of two-level 2D CCWT. (a) is a color image, (b) is the low-frequency sub-band of (a), and (c)-(f) are the White-Black, Red-Cyan, Green-Magenta and Blue-Yellow complementary color sub-bands of (a), respectively.

image, (b) is the low-frequency sub-band of (a), and (c)-(f) are the White-Black, Red-Cyan, Green-Magenta and Blue-Yellow complementary color sub-bands of (a), respectively.

### B. Low-Frequency Feature Extraction Via Block-Based CS

The block-based CS is utilized to extract low-frequency feature from the low-frequency sub-band. The CS theory was proposed by Donoho [36]. The CS makes a breakthrough from the limitations of the Nyquist sampling theorem and enables compression coding during sampling. It has been successfully used in various fields, such as audio, holography and MRI [37]–[39]. Specifically, CS theory shows that as long as the signal is sparse in the transform domain, it can be measured by an observation matrix unrelated to transform bases and the original signal can be reconstructed from a small amount of sampling data. Let  $\mathbf{X} \in \mathbb{R}^{N \times N}$  be the test image and  $\mathbf{w}$  be a wavelet orthogonal transform matrix sized  $N \times N$ . Thus, the sparse representation  $\mathbf{X}'$  of  $\mathbf{X}$  is calculated by the below formula [40].

$$\mathbf{X}' = \mathbf{w}\mathbf{X}\mathbf{w}^\Delta \quad (6)$$

where  $\Delta$  is the transposition-conjugate operator. Next,  $\mathbf{X}'$  is measured by an observation matrix  $\Psi \in \mathbb{R}^{M \times N}$  ( $M < N$ ). Then, a measurement matrix  $\mathbf{Y}$  with size  $M \times N$  can be represented as follows:

$$\mathbf{Y} = \Psi\mathbf{X}' \quad (7)$$

Here, we utilize DWT to generate sparse representation  $\mathbf{X}'$  and use Gaussian matrix as the observation matrix. Moreover, the compression rate  $M/N$  is selected as 0.5. This is based on the consideration that the high compression rate cannot show the advantages of the low sampling rate of CS, while the low compression rate increases the possibility of losing important information.

To extract low-frequency feature via block-based CS, we divide the low-frequency sub-band  $\mathbf{P}^{(L)}$  into non-overlapping blocks with size  $b \times b$ . Let  $U_l$  be the integral multiple of  $b$ . Thus, there are a total of  $F = (U_l/b)^2$  blocks.

Suppose that  $\mathbf{B}_i$  ( $1 \leq i \leq F$ ) is the  $i$ -th block. To further calculate the relevance of information in the block, we divide every block into two parts sized  $b \times b/2$ , i.e., a left sub-block and a right sub-block. To conduct CS, the two sub-blocks are then converted to a squared size  $b \times b$  by bi-cubic interpolation. Let the resized left part and right part of  $\mathbf{B}_i$  be  $\mathbf{B}_i^{(1)}$  and  $\mathbf{B}_i^{(2)}$ , respectively. We apply CS to the two parts, and then obtain two measurement matrices  $\mathbf{Y}_i^{(1)}$  and  $\mathbf{Y}_i^{(2)}$  as follows:

$$\mathbf{Y}_i^{(1)} = \Psi' \mathbf{B}_i^{(1)'} \quad (8)$$

$$\mathbf{Y}_i^{(2)} = \Psi' \mathbf{B}_i^{(2)'} \quad (9)$$

where  $\mathbf{B}_i^{(1)'}$  and  $\mathbf{B}_i^{(2)'}$  are the sparse representations of  $\mathbf{B}_i^{(1)}$  and  $\mathbf{B}_i^{(2)}$  obtained by DWT,  $\Psi'$  is the observation matrix sized  $b/2 \times b$ , and the sizes of  $\mathbf{Y}_i^{(1)}$  and  $\mathbf{Y}_i^{(2)}$  are both  $b/2 \times b$ .

To make the extracted features reflect the fluctuation of measurement matrix elements, we select variance as the CS feature. The variances of the two measurement matrices can be computed by the following formulas.

$$e_i^{(1)} = \frac{\sum_{u=1}^b \sum_{v=1}^{b/2} [Y_i^{(1)}(u, v) - \mu_{Y_i^{(1)}}]^2}{S - 1} \quad (10)$$

$$e_i^{(2)} = \frac{\sum_{u=1}^b \sum_{v=1}^{b/2} [Y_i^{(2)}(u, v) - \mu_{Y_i^{(2)}}]^2}{S - 1} \quad (11)$$

where  $S = b^2/2$  is the element numbers of the  $\mathbf{Y}_i^{(1)}$  and  $\mathbf{Y}_i^{(2)}$ ,  $Y_i^{(1)}(u, v)$  and  $Y_i^{(2)}(u, v)$  are the elements in the  $u$ -th row and the  $v$ -th column of  $\mathbf{Y}_i^{(1)}$  and  $\mathbf{Y}_i^{(2)}$ ,  $\mu_{Y_i^{(1)}}$  and  $\mu_{Y_i^{(2)}}$  are the means of  $\mathbf{Y}_i^{(1)}$  and  $\mathbf{Y}_i^{(2)}$ , respectively. They are defined as follows:

$$\mu_{Y_i^{(1)}} = \frac{\sum_{u=1}^b \sum_{v=1}^{b/2} Y_i^{(1)}(u, v)}{S} \quad (12)$$

$$\mu_{Y_i^{(2)}} = \frac{\sum_{u=1}^b \sum_{v=1}^{b/2} Y_i^{(2)}(u, v)}{S} \quad (13)$$

Next, we concatenate the CS features of all left and right sub-blocks, respectively. The block-based CS feature vectors are available by the below formulas.

$$\mathbf{e}^{(1)} = [e_1^{(1)}, e_2^{(1)}, \dots, e_F^{(1)}] \quad (14)$$

$$\mathbf{e}^{(2)} = [e_1^{(2)}, e_2^{(2)}, \dots, e_F^{(2)}] \quad (15)$$

Considering that two CS features are extracted from every block, we take them as a point, i.e.,  $(e_i^{(1)}, e_i^{(2)})$ , and compress features by computing the Euclidean distance from the point to a reference point. To do so, a reference point  $(\mu_{e^{(1)}}, \mu_{e^{(2)}})$  is constructed, where  $\mu_{e^{(1)}}$  and  $\mu_{e^{(2)}}$  are the means of  $\mathbf{e}^{(1)}$  and  $\mathbf{e}^{(2)}$ , respectively. Thus, the reference point is computed by the below equations:

$$\mu_{e^{(1)}} = \frac{1}{F} \sum_{i=1}^F e_i^{(1)} \quad (16)$$

$$\mu_{e^{(2)}} = \frac{1}{F} \sum_{i=1}^F e_i^{(2)} \quad (17)$$

The Euclidean distance from the point  $(e_i^{(1)}, e_i^{(2)})$  to the reference point  $(\mu_{e^{(1)}}, \mu_{e^{(2)}})$  is calculated as follows:

$$d_i = \sqrt{(e_i^{(1)} - \mu_{e^{(1)}})^2 + (e_i^{(2)} - \mu_{e^{(2)}})^2} \quad (18)$$

Finally, the low-frequency feature sequence  $\mathbf{D}_1$  are available by concatenating  $d_i$  as follows:

$$\mathbf{D}_1 = [d_1, d_2, \dots, d_F] \quad (19)$$

The low-frequency feature sequence  $\mathbf{D}_1$  has  $F$  floating-point numbers.

### C. Multi-Directional Edge Detection

As an important visual feature, image edge is the boundary of the area with significant local variations in brightness. It has been widely used in many applications of computer vision and image processing [41]–[43]. Since image edge can represent textures of image, numerous studies use edge detection methods to improve sensitivity to image distortion. Generally, image edge detection methods can find edge information via gradient algorithm and thus the extracted edges contain clear image contours and rich textures. As the complementary color sub-bands reflect multi-directional color information, the color features extracted from the edge of these sub-bands can describe image distortion on different color channels. Here, we use the multi-directional Sobel operator [44] to extract image edge from complementary color sub-bands. Our considerations of choosing the multi-directional Sobel operator for edge detection are as follows. The multi-directional Sobel operator is an improved technique of the well-known Sobel operator by using eight directions. The edge detection with the multi-directional Sobel operator is more accurate than the edge detection with the classic Sobel operator. In fact, the multi-directional Sobel operator can reach a good balance between computational cost and accuracy of edge detection. Moreover, the eight directions of the multi-directional Sobel operator, i.e.,  $n = \frac{m\pi}{8}$  ( $m = 1, 2, \dots, 8$ ), coincide with the directions of our complementary color sub-bands, which contributes to detecting rich edge information. Specifically, through eight  $5 \times 5$  convolution templates, the gradients in eight directions of each complementary color sub-band can be calculated by the equations (20)–(27):

$$\mathbf{G}_{\frac{\pi}{8}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & -4 & -2 & 0 \\ -1 & -4 & 0 & 4 & 1 \\ 0 & 2 & 4 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} * \mathbf{f} \quad (20)$$

$$\mathbf{G}_{\frac{\pi}{4}} = \begin{bmatrix} 0 & 0 & 0 & -1 & 0 \\ 0 & -2 & -4 & 0 & 1 \\ 0 & -4 & 0 & 4 & 0 \\ -1 & 0 & 4 & 2 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} * \mathbf{f} \quad (21)$$

$$\mathbf{G}_{\frac{3\pi}{8}} = \begin{bmatrix} 0 & 0 & -1 & 0 & 0 \\ 0 & -2 & -4 & 2 & 0 \\ 0 & -4 & 0 & 4 & 0 \\ 0 & -2 & 4 & 2 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} * \mathbf{f} \quad (22)$$

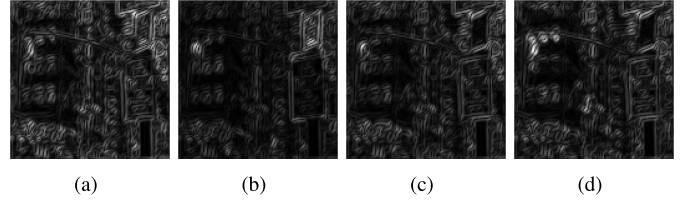


Fig. 4. Visual examples of the maximum gradient maps corresponding to different complementary color sub-bands. (a)–(d) are the corresponding maximum gradient maps of Figure 3 (c)–(f).

$$\mathbf{G}_{\frac{\pi}{2}} = \begin{bmatrix} 0 & -1 & 0 & 1 & 0 \\ 0 & -2 & 0 & 2 & 0 \\ 0 & -4 & 0 & 4 & 0 \\ 0 & -2 & 0 & 2 & 0 \\ 0 & -1 & 0 & 1 & 0 \end{bmatrix} * \mathbf{f} \quad (23)$$

$$\mathbf{G}_{\frac{5\pi}{8}} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & -2 & 4 & 2 & 0 \\ 0 & -4 & 0 & 4 & 0 \\ 0 & -2 & -4 & 2 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix} * \mathbf{f} \quad (24)$$

$$\mathbf{G}_{\frac{3\pi}{4}} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 4 & 2 & 0 \\ 0 & -4 & 0 & 4 & 0 \\ 0 & -2 & -4 & 0 & 1 \\ 0 & 0 & 0 & -1 & 0 \end{bmatrix} * \mathbf{f} \quad (25)$$

$$\mathbf{G}_{\frac{7\pi}{8}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 4 & 2 & 0 \\ -1 & -4 & 0 & 4 & 1 \\ 0 & -2 & -4 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} * \mathbf{f} \quad (26)$$

$$\mathbf{G}_{\pi} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ -1 & -2 & -4 & -2 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 4 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} * \mathbf{f} \quad (27)$$

in which  $*$  is the convolution operation, and  $\mathbf{f}$  represents the complementary color sub-band. For every complementary color sub-band, we compare the absolute values of the gradients in the eight convolution results, and then output the maximum value to generate the maximum gradient map  $\mathbf{G}$ . Let  $G(p, q)$  be the element of  $\mathbf{G}$  at the position  $(p, q)$ . It can be denoted as follows:

$$\begin{aligned} G(p, q) &= \max \left\{ \left| G_{\frac{\pi}{8}}(p, q) \right|, \left| G_{\frac{\pi}{4}}(p, q) \right|, \left| G_{\frac{3\pi}{8}}(p, q) \right|, \left| G_{\frac{\pi}{2}}(p, q) \right|, \right. \\ &\quad \left. \left| G_{\frac{5\pi}{8}}(p, q) \right|, \left| G_{\frac{3\pi}{4}}(p, q) \right|, \left| G_{\frac{7\pi}{8}}(p, q) \right|, \left| G_{\pi}(p, q) \right| \right\} \quad (28) \end{aligned}$$

Thus, the maximum gradient maps of  $\mathbf{P}_{\text{total}}^{(\text{W-B})}$ ,  $\mathbf{P}_{\text{total}}^{(\text{R-C})}$ ,  $\mathbf{P}_{\text{total}}^{(\text{G-M})}$  and  $\mathbf{P}_{\text{total}}^{(\text{B-Y})}$  can be generated by similar calculations. Figure 4 demonstrates visual examples of the maximum gradient maps corresponding to different complementary color sub-bands.

#### D. Color Feature Extraction Via Block-Based CS

To extract color features from the maximum gradient maps, the block-based CS described in Section III-B is adopted again. The four complementary color features are extracted from the corresponding maximum gradient maps via the block-based CS, and then they are concatenated to generate the final color features. Here, take the White-Black feature extraction for example. Firstly, we divide the maximum gradient map of  $\mathbf{P}_{\text{total}}^{(W-B)}$  into non-overlapping blocks with  $k \times k$  size. Let  $U_l$  be the integral multiple of  $k$ . Thus, there are a total of  $E = (U_l/k)^2$  blocks. Suppose that  $\mathbf{K}_j$  ( $1 \leq j \leq E$ ) is the  $j$ -th block. Then, every block is divided into two sub-blocks, the two sub-blocks are resized, and CS is applied to them. Next, the block-based CS features are taken as a point and the Euclidean distance from the point to a reference point is computed. Finally, the White-Black feature sequence  $\mathbf{D}^{(W-B)}$  is available by concatenating all distances. Likewise, the Red-Cyan feature sequence  $\mathbf{D}^{(R-C)}$ , the Green-Magenta feature sequence  $\mathbf{D}^{(G-M)}$  and the Blue-Yellow feature sequence  $\mathbf{D}^{(B-Y)}$  can be extracted from the corresponding maximum gradient maps, respectively. Therefore, the final color features are available by concatenating four complementary color feature sequences as follows:

$$\mathbf{D}_2 = [\mathbf{D}^{(W-B)}, \mathbf{D}^{(R-C)}, \mathbf{D}^{(G-M)}, \mathbf{D}^{(B-Y)}] \quad (29)$$

The color features  $\mathbf{D}_2$  consist of  $4E$  floating-point numbers.

#### E. Hash Construction

The extracted perceptual features from the CCWT sub-bands, i.e., the low-frequency feature sequence  $\mathbf{D}_1$  and the color feature sequence  $\mathbf{D}_2$ , are combined to construct the initial hash sequence as follows:

$$\mathbf{D} = [\alpha \mathbf{D}_1, \beta \mathbf{D}_2] \quad (30)$$

where the parameters  $\alpha$  and  $\beta$  are used to adjust the influence of  $\mathbf{D}_1$  and  $\mathbf{D}_2$ . Clearly, the length of  $\mathbf{D}$  is  $Z = F + 4E$ . Let  $d(t)$  be the  $t$ -th element of the sequence  $\mathbf{D}$  ( $1 \leq t \leq Z$ ). Since  $d(t)$  is a floating-point number,  $d(t)$  is quantized to an integer for reducing storage by the below equation.

$$h(t) = \text{round}[d(t) + 0.5] \quad (31)$$

where  $\text{round}[\cdot]$  is the rounding operation. The final hash is obtained as follows:

$$\mathbf{h} = [h(1), h(2), \dots, h(Z)] \quad (32)$$

Consequently, the final hash  $\mathbf{h}$  is a sequence of  $Z$  integers.

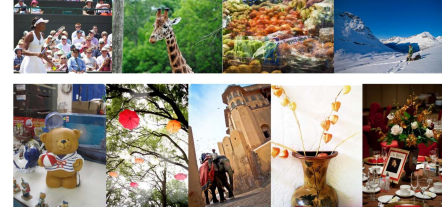
#### F. Hash Similarity Evaluation

We take the  $L_2$  norm to measure similarity of two image hashes. Suppose that  $\mathbf{h}_1 = [h_1(1), h_1(2), \dots, h_1(Z)]$  and  $\mathbf{h}_2 = [h_2(1), h_2(2), \dots, h_2(Z)]$  are hash sequences of two images. The  $L_2$  norm of the two hash sequences is defined by the following equation:

$$d_{\text{norm}}(\mathbf{h}_1, \mathbf{h}_2) = \sqrt{\sum_{t=1}^Z [h_1(t) - h_2(t)]^2} \quad (33)$$



(a) UCID



(b) Test set of COCO 2017 database

Fig. 5. Sample images from the used databases.

where  $h_1(t)$  and  $h_2(t)$  are the  $t$ -th elements of  $\mathbf{h}_1$  and  $\mathbf{h}_2$ , respectively. Generally, a smaller  $d_{\text{norm}}$  means more similar images of the input hash sequences. Therefore, for different images, their corresponding  $d_{\text{norm}}$  should be a big value.

## IV. EXPERIMENTAL RESULTS

### A. Experimental Settings

The parameters of the proposed hashing algorithm are set as follows. Pre-processing operations are exploited to convert the size of input image to  $512 \times 512$ , the low-frequency sub-band is divided into blocks with the size  $32 \times 32$ , the complementary color sub-bands are divided into blocks with the size  $32 \times 32$ , and the weight parameters are 0.5 and 2.5, i.e.,  $U = 512$ ,  $b = 32$ ,  $k = 32$ ,  $\alpha = 0.5$  and  $\beta = 2.5$ . Therefore,  $Z = 80$ . Section IV-E and Section IV-F discuss the effect of block size selection and the effect of weight parameter selection, respectively. Besides, two databases are selected for comprehensive experiments. Experiments are carried out on a computer with a 2.90 GHz Intel Core i5-10400 CPU and 16.0 GB RAM. The coding platform is the MATLAB R2019b and the adopted OS is Windows 10. Each of the used databases is detailed below, where some sample images from the databases are shown in Figure 5.

1) *UCID-Based Database*: It consists of 100350 color images for robustness analysis, in which image sizes are  $512 \times 384$  or  $384 \times 512$ . Specifically, the open database UCID [45] comprises 1338 color images. Ten types of content-preserving operations are applied to 1338 color images by MATLAB and StirMark [46] for generating similar images. These operations include contrast adjustment (CA), brightness adjustment (BA), gamma correction (GC), speckle noise (SN),  $3 \times 3$  Gaussian low-pass filtering (GLF) and salt and pepper noise (SPN) which are all implemented by MATLAB, and watermark embedding (WE), image scaling (IS), JPEG compression (JC) and the combinational operation (CO) of rotation, cropping and rescaling which are achieved by StirMark. Furthermore, the detailed parameter values of these specific operations are as follows. The used magnitudes of CA and BA are

$\pm 20$  and  $\pm 10$ . The selected parameter values of GC are 1.25, 1.1, 0.9 and 0.75. The used variance range of SN is from 0.001 to 0.01 in the step of 0.001. The used standard deviation range of GLF is from 0.3 to 1.0 in the step of 0.1. The used density range of SPN is from 0.001 to 0.01 in the step of 0.001. The used strength range of WE is from 10 to 100 in the step of 10. The selected ratios of IS are 2.0, 1.5, 1.1, 0.9, 0.75 and 0.5. The used quality factor range of JC is from 30 to 100 in the step of 10. The used rotation angles of CO are  $\pm 5^\circ$ ,  $\pm 4^\circ$ ,  $\pm 3^\circ$ ,  $\pm 2^\circ$  and  $\pm 1^\circ$ . As there are 74 operations in total, 74 similar versions generated from each color image are obtained. Consequently, there are  $74 \times 1338 = 99012$  pairs of similar images and a total of  $99012 + 1338 = 100350$  images for robustness analysis.

2) *Test Set of COCO 2017 Database*: This dataset [47] comprises 40640 color images for discrimination evaluation. It can be observed that the color images in this database contain a variety of contents, such as human beings, animals, buildings, sports, natural scenes, and man-made objects. This database can generate 40640 hash sequences of different images. Thus, hash similarity is computed and then  $C_{40640}^2 = 40640 \times (40640 - 1)/2 = 825784480$   $L_2$  norms in total are obtained.

3) *Evaluation Criterion*: To effectively evaluate the classification performance of our hashing algorithm in terms of robustness and discrimination, we adopt the famed tool called ROC graph [48] as the evaluation criterion. In the ROC graph, false positive rate and true positive rate are viewed as a set of points with coordinates for plotting the ROC curve. For hashing algorithms, the false positive rate indicates discrimination and the true positive rate represents robustness. The equations of the true positive rate  $R_t$  and the false positive rate  $R_f$  are given as follows:

$$R_t = \frac{n_{\text{correct}}}{N_{\text{similar}}} \quad (34)$$

$$R_f = \frac{n_{\text{wrong}}}{N_{\text{different}}} \quad (35)$$

in which  $n_{\text{correct}}$  is the number of correctly identified similar images,  $n_{\text{wrong}}$  is the number of wrongly distinguished different images,  $N_{\text{similar}}$  and  $N_{\text{different}}$  are the number of similar images and different images, respectively. Clearly, good performance of a hashing algorithm means a high  $R_t$  and a low  $R_f$ . This implies that in the ROC graph, the curve closer to the top-left corner has better performance. For the Area Under the Curve (AUC), its value range is  $[0, 1]$ . Therefore, the larger the AUC, the higher the classification accuracy.

### B. Robustness Analysis

To analyse the robustness of our hashing algorithm, the experiment is carried out on the UCID-based database. In detail, we evaluate hash similarity between 99012 pairs of similar images with  $L_2$  norm. Generally, small  $L_2$  norms of similar image pairs demonstrate that the hashing algorithm can correctly identify similar images. The mean  $L_2$  norms under different parameters of content-preserving operations are displayed in Figure 6. The  $x$ -axis of the figure represents the used parameter of the specific operations, and the  $y$ -axis of the

TABLE II  
STATISTICAL RESULTS OF  $L_2$  NORM UNDER DIFFERENT OPERATIONS

Operation	Maximum	Minimum	Mean	Standard deviation
CA	62.88	2.00	12.23	7.80
BA	117.06	6.93	38.38	15.95
SN	40.84	1.00	8.18	4.20
SPN	29.00	1.41	6.01	2.58
GC	74.62	5.20	22.64	11.90
GLF	83.96	0	14.98	12.69
JC	87.45	5.66	15.64	5.75
WE	62.82	1.00	12.83	8.20
IS	63.84	4.90	12.90	4.55
CO	266.75	9.54	54.38	24.96

figure indicates the mean  $L_2$  norm. It can be noticed that the mean  $L_2$  norms under different operations are smaller than 60, except the CO as shown in Figure 6 (j). In addition, it is found that the mean  $L_2$  norms are small and change slightly in Figures 6 (c), (d), (g) and (i). These results demonstrate that our hashing algorithm is robust enough against SN, SPN, JC and IS. The mean  $L_2$  norm of CO is larger than those of other operations because the CO introduces multiple distortions. All averages of the mean  $L_2$  norms are smaller than 60, indicating that our hashing algorithm can resist content-preserving operations.

Furthermore, we tabulate the maximum, minimum, mean and standard deviation of  $L_2$  norm under all operations in Table II. It is clearly seen that the means of all operations are smaller than 40, except the CO. In addition, most statistical results of  $L_2$  norm are small, implying that the proposed hashing algorithm can effectively identify similar images. To further understand the identification accuracy, we calculate the correct detection rate of similar images under different thresholds. Specifically, when the threshold  $T = 50$ , 91.37% similar images are accurately identified. And our hashing algorithm can successfully identify 98.77% or 99.96% similar images when  $T = 90$  or  $T = 150$ , respectively. These results illustrate that our hashing algorithm can maintain a high correct detection rate which satisfies the need for excellent robustness.

### C. Discrimination Evaluation

The test set of COCO 2017 database is adopted to evaluate discrimination performance of our hashing algorithm. Similarly, we evaluate hash similarity between 825784480 pairs of different images with  $L_2$  norm. The distribution of all  $L_2$  norms is depicted in Figure 7, where the  $x$ -axis of this figure is the  $L_2$  norm and the  $y$ -axis of this figure is the corresponding frequency. The statistical results show that the minimum and maximum of all  $L_2$  norms are 22.83 and 879.83, respectively. The standard deviation and mean of all  $L_2$  norms are 50.68 and 214.37, respectively. It can be seen that there is a great difference between the mean  $L_2$  norm of different images and the mean  $L_2$  norm of similar images, which demonstrates that our hashing algorithm has good discrimination. However, the minimum  $L_2$  norm of different images is smaller than the maximum  $L_2$  norm of similar images. This



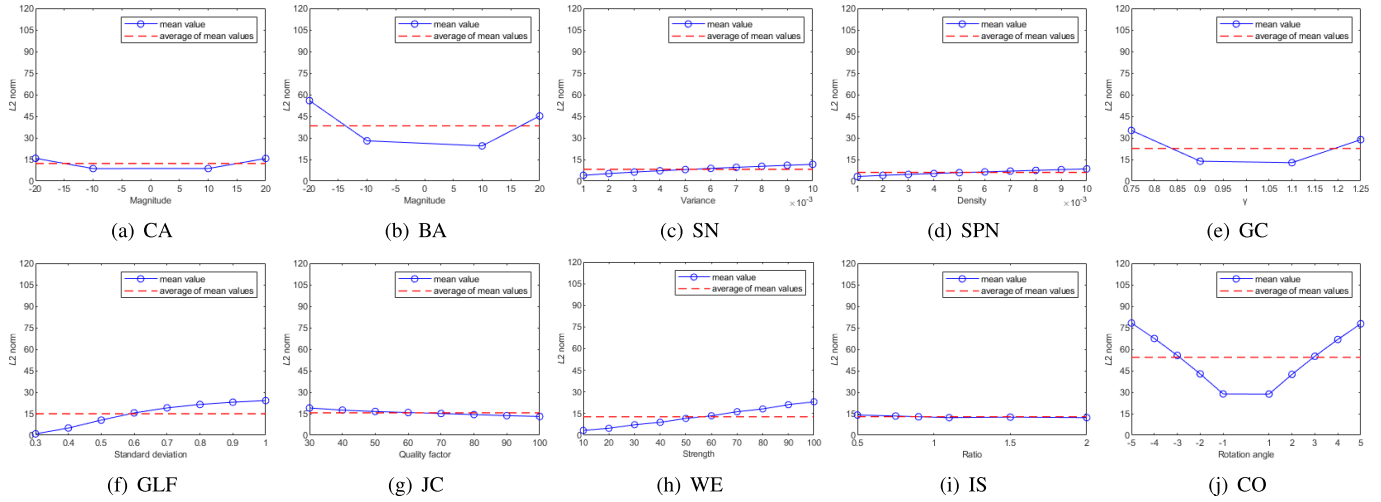


Fig. 6.  $L_2$  norm results under content-preserving operations.

TABLE III  
DETECTION PERFORMANCES UNDER DIFFERENT THRESHOLDS

Threshold	CDR (%)	FDR (%)
150	99.96	6.95
130	99.89	1.90
110	99.62	0.33
90	98.77	0.03048
70	96.53	0.0012454
50	91.37	0.00001986

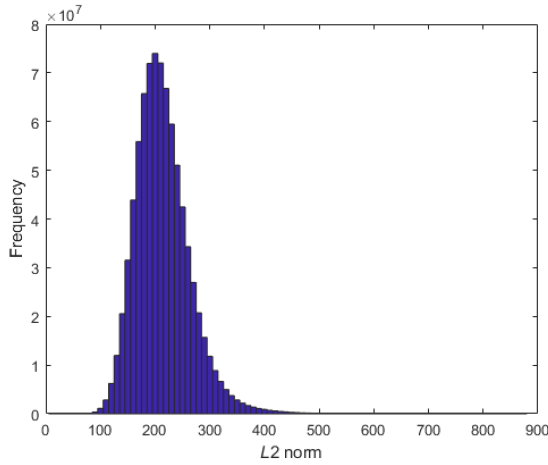


Fig. 7. Distribution of 825784480 results of  $L_2$  norm.

indicates that there is a certain classification error between different images and similar images. The correct detection rate (CDR) and false detection rate (FDR) of our hashing algorithm under different thresholds are tabulated in Table III, where CDR is the indicator of robustness, and FDR is the indicator of discrimination. Obviously, FDR decreases with the decrease of the threshold. A small threshold means a low probability of classifying different images into similar ones, indicating good discrimination. However, since robustness and discrimination are mutually constrained, CDR also decreases.

Therefore, a proper threshold should be selected to balance the discrimination and robustness according to the practical application.

#### D. Hash Storage

To calculate the cost of storing a hash sequence, we use the test set of COCO 2017 database in the discrimination experiment as the data source. As our hash sequence consists of 80 integers, the total number of  $80 \times 40640 = 3251200$  elements is generated from 40640 color images by our hashing algorithm. Statistical results of 3251200 hash elements are as follows. The minimum and maximum element values are 1 and 481, respectively. Obviously, 9 bits can represent an integer in the range of  $0$  to  $2^9 - 1 = 511$ . Consequently, one hash element storage requires 9 bits, and the storage cost of our hash is  $80 \times 9 = 720$  bits in binary form.

#### E. Block Size Selection

Considering the importance of the block-based CS in our hashing algorithm, we discuss the effect of different block sizes on classification performances. Note that  $b$  represents block size on the low-frequency sub-band and  $k$  represents block size on the maximum gradient maps. Here, we only change  $b$  and  $k$  and keep other parameters constant. The used block sizes are selected from  $b \in \{16, 32, 64\}$  and  $k \in \{32, 64\}$ . We do not select  $k = 16$  because the hash length is too long under this parameter value. Therefore, there are six combinations of the block sizes as follows: (1)  $b = 16, k = 32$ ; (2)  $b = 16, k = 64$ ; (3)  $b = 32, k = 32$ ; (4)  $b = 32, k = 64$ ; (5)  $b = 64, k = 32$ ; (6)  $b = 64, k = 64$ . The hash lengths of these six combinations are 128, 80, 80, 32, 68 and 20, respectively. The UCID-based database and the test set of COCO 2017 database are taken to generate ROC curves, i.e., the calculation of  $R_t$  and  $R_f$  under a set of thresholds by using 99012 pairs of similar images and 825784480 pairs of different images. Figure 8 illustrates ROC curves under different combinations of block sizes. It is observed that the first two curves close to the top-left corner

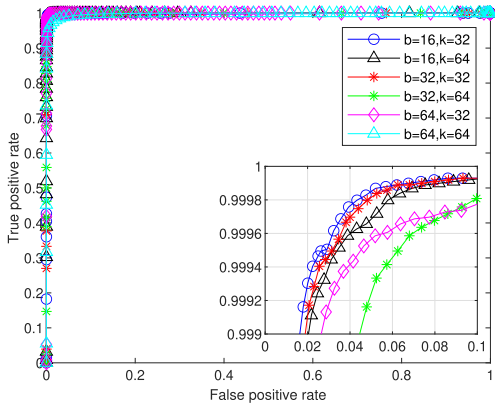


Fig. 8. ROC curves under different combinations of block sizes.

are almost overlapping, i.e.,  $b = 16$ ,  $k = 32$  and  $b = 32$ ,  $k = 32$ , meaning that they have similar good classification performance. Nevertheless, for the storage cost, the hash length of  $b = 32$  and  $k = 32$  is shorter than that of  $b = 16$  and  $k = 32$ . Taken together, when  $b = 32$  and  $k = 32$ , the best performance of the proposed algorithm can be achieved in terms of classification and hash storage.

#### F. Selection of Weight Parameters

Apart from the block size, the weight parameters are another important factors influencing the performance of our hashing algorithm. Note that the final hash sequence consists of two main components, where  $\mathbf{D}_1$  represents low-frequency features and  $\mathbf{D}_2$  indicates color features. The weight parameters  $\alpha$  and  $\beta$  are used to adjust the contributions of  $\mathbf{D}_1$  and  $\mathbf{D}_2$  in classification performance. We still only change  $\alpha$  and  $\beta$  and keep other parameters constant.

Considering that the influences of content-preserving operations mainly concentrate on high-frequency sub-bands and the changes of low-frequency sub-bands are slight, we reduce the importance of  $\mathbf{D}_1$  and enhance the importance of  $\mathbf{D}_2$  for balancing their contributions. Specifically, the values adopted for  $\alpha$  are 0, 0.3, 0.5, 0.7 and 1, and the values adopted for  $\beta$  are 1, 1.5, 2, 2.5 and 3. The AUCs of our hashing algorithm with different values of  $\alpha$  and  $\beta$  are calculated for evaluating classification performance. The results are displayed in Table IV. To make easy comparison, the AUC results under different weight parameters are visually plotted in the Figure 9. Notice that the parameter selection of  $\alpha = 0$  and  $\beta = 1$  means that the final hash is only composed of color features. For  $\alpha = 1$ , as the  $\beta$  value increases, the AUC also increases and reaches the maximum value when  $\beta = 2.0$  or  $\beta = 2.5$ . This indicates that color features have a positive effect on classification performance. Overall, all AUCs are bigger than 0.99900. Therefore, we can select suitable weight parameters according to the specific performance requirement of the application. In this paper, the target application of our hashing algorithm is RR IQA. Since  $\mathbf{D}_2$  helps to improve perceptual sensitivity, the best result is generated with  $\alpha = 0.5$  and  $\beta = 2.5$  for the application of RR IQA.

TABLE IV  
AUC COMPARISONS UNDER DIFFERENT WEIGHT PARAMETERS

$\beta \backslash \alpha$	0	0.3	0.5	0.7	1.0
1.0	0.99932	0.99993	0.99996	0.99994	0.99990
1.5	0.99933	0.99986	0.99995	0.99996	0.99995
2.0	0.99933	0.99978	0.99991	0.99995	0.99996
2.5	0.99933	0.99971	0.99986	0.99993	0.99996
3.0	0.99933	0.99964	0.99981	0.99989	0.99993

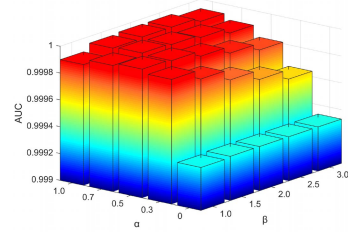


Fig. 9. Visual results of AUC under different weight parameters.

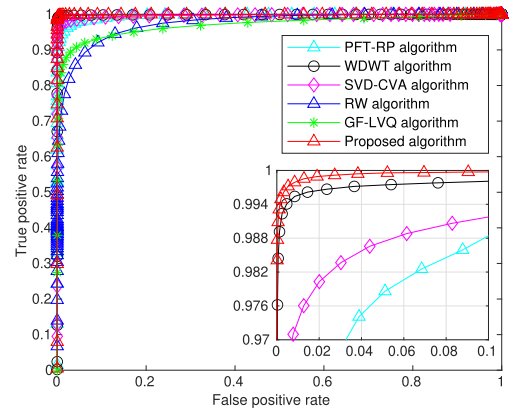


Fig. 10. ROC performance comparison among different algorithms.

#### G. Performance Comparison Among Different Algorithms

The superior performance of our hashing algorithm is demonstrated by making comparisons with some advanced hashing algorithms, such as PFT-RP algorithm [22], WDWT algorithm [5], SVD-CVA algorithm [27], RW algorithm [30] and GF-LVQ algorithm [29]. The UCID-based database and the test set of COCO 2017 database are deployed to validate performances. All images are converted to  $512 \times 512$  before they are input to the compared hashing algorithms. Here the parameter settings and similarity metrics of these compared algorithms are consistent with those of their original papers. More specifically, for the similarity metric, PFT-RP algorithm adopts the  $L_1$  norm, WDWT algorithm and SVD-CVA algorithm adopt the  $L_2$  norm, RW algorithm and GF-LVQ algorithm adopt the normalized Hamming distance. The experimental results of our hashing algorithm with  $b = 32$ ,  $k = 32$ ,  $\alpha = 0.5$  and  $\beta = 2.5$  are chosen for comparison.

Classification performances of different hashing algorithms are demonstrated in an ROC graph. Specifically, the ROC

TABLE V  
ALGORITHM PERFORMANCE COMPARISON

Algorithm	AUC	Time (s)	Length (bit)
PFT-RP [22]	0.99623	0.07	459
WDWT [5]	0.99920	0.35	640
SVD-CVA [27]	0.99697	8.70	3328
RW [30]	0.97500	0.05	144
GF-LVQ [29]	0.97048	0.58	120
Proposed	0.99986	0.22	720

curves of the proposed hashing algorithm and the compared algorithms are all displayed in Figure 10. According to this figure, we can observe that the curves of all algorithms are near the top-left corner of the ROC graph, and our curve is closer than those of other algorithms. Table V lists the AUC values of different algorithms. The AUC values of the proposed algorithm, PFT-RP algorithm, WDWT algorithm, SVD-CVA algorithm, RW algorithm and GF-LVQ algorithm are 0.99986, 0.99623, 0.99920, 0.99697, 0.97500 and 0.97048, respectively. Obviously, the AUC value of the proposed algorithm is the largest one, demonstrating that our hashing algorithm has better classification performance than other algorithms. Our hashing algorithm can achieve better classification performance. This is mainly contributed by the uses of CCWT and CS. The calculation of CCWT exploits all color channels without discarding any information. Since the low-frequency sub-bands of CCWT contain basic image information, the features extracted from these sub-bands can distinguish images with different contents. In addition, CS can directly achieve compression coding during the sampling process. As the Euclidean distance between the block-based CS features is slightly influenced by content-preserving operations, perceptual features constructed by Euclidean distances are robust, discriminative and compact.

Moreover, the computational time and hash length of these hashing algorithm are reported in Table V, where the computational time is measured by the mean time of producing a hash sequence. The computational time of the proposed algorithm, PFT-RP algorithm, WDWT algorithm, SVD-CVA algorithm, RW algorithm and GF-LVQ algorithm are 0.22, 0.07, 0.35, 8.70, 0.05 and 0.58 seconds, respectively. The proposed algorithm is faster than WDWT algorithm, SVD-CVA algorithm and GF-LVQ algorithm. The hash lengths of the proposed algorithm, PFT-RP algorithm, WDWT algorithm, SVD-CVA algorithm, RW algorithm and GF-LVQ algorithm are 720, 459, 640, 3328, 144 and 120 bits. Our length is much shorter than that of SVD-CVA algorithm, but it is longer than those of other algorithms.

## V. APPLICATION IN REDUCED-REFERENCE IMAGE QUALITY ASSESSMENT

In many practical applications, when similar versions of the original image are correctly identified, the assessment of perceptual image quality is then needed. IQA is an increasingly significant issue [49], [50]. RR IQA is a branch of IQA and it assesses the distortion level of an image by extracting its content representation. Therefore, RR IQA requires three conditions as follows: (i) Provide a content-based image

representation; (ii) Perceive the level of distortion on similar images; (iii) Predict score relevant to the visual quality of image. Coincidentally, perceptual image hashing satisfies these conditions. It can produce a visual content-based sequence from image and has the ability of sensitivity for measuring various image distortions. Thus, perceptual image hashing can be used for RR IQA. When our hashing algorithm is applied to the RR IQA, the specific process is as follows. At the sender's side, the hash of the reference image is generated by our hashing algorithm and then it is sent to the receiver via auxiliary channel. Meanwhile, the reference image is also sent to the receiver through transmission channel. At the receiver's side, the distorted version of the reference image and the hash of the reference image are both received. Next, the hash of the distorted image is extracted by our hashing algorithm. Finally, objective score is obtained by calculating the  $L_2$  norm between the received hash and the extracted hash. To validate the effectiveness of our hashing algorithm in RR IQA, two databases are adopted to compare the IQA performance with different schemes in terms of some well-known evaluation criteria.

### A. Experimental Databases and Evaluation Criteria

1) *Experimental Databases*: Two image databases are used to verify our application in RR IQA. The first database called LIVE [51] consists of 779 distorted versions of 29 reference images with 5 distortion types. These distorted versions contain 145 white noisy images, 175 JPEG compressed images, 169 JPEG2000 compressed images, 145 Gaussian blurred images and 145 fast-fading Rayleigh channel noisy images. The LIVE database provides the differential mean opinion score (DMOS) of subjective score. The second database called TID2013 [52] consists of 25 reference images and its 3000 distorted versions with 24 distortion types. The TID2013 database provides the mean opinion score (MOS) of subjective score.

2) *Evaluation Criteria*: Four objective performance evaluation criteria are adopted to evaluate the correlation between objective predicted scores and subjective scores, i.e., Spearman Rank-order Correlation Coefficient (SRCC), Pearson Linear Correlation Coefficient (PLCC), Kendalls Rank-order Correlation Coefficient (KRCC) and Root Mean Squared Error (RMSE), where PLCC and RMSE indicate accuracy and consistency of predictions, respectively, and SRCC and KRCC reflect monotonicity. Let  $W$  be the number of distorted images,  $x_i$  be the converted objective predicted score of the  $i$ -th image after nonlinear regression, and  $y_i$  be the subjective score of the  $i$ -th image. Then, these performance criteria are defined as follows:

$$\text{SRCC} = 1 - \frac{6 \sum_{i=1}^W s_i^2}{W(W^2 - 1)} \quad (36)$$

$$\text{PLCC} = \frac{\sum_{i=1}^W (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^W (x_i - \mu_x)^2 (y_i - \mu_y)^2}} \quad (37)$$

$$\text{KRCC} = \frac{2(W_{\text{con}} - W_{\text{dis}})}{W(W - 1)} \quad (38)$$

TABLE VI  
IQA PERFORMANCE COMPARISONS ON DIFFERENT DATABASES

Database	Criterion	FR IQA		RR IQA				
		PSNR	SSIM	WDWT	FSI	OSVP	WNISM	Proposed
LIVE (779)	SRCC	0.8756	<b>0.9479</b>	0.8821	0.8826	0.8264	0.7472	<b>0.9006</b>
	PLCC	0.8723	<b>0.9449</b>	0.8771	0.8821	0.8265	0.7360	<b>0.8998</b>
	KRCC	0.6865	<b>0.7963</b>	0.6881	0.6957	0.6330	0.5576	<b>0.7231</b>
	RMSE	13.3597	<b>8.9455</b>	13.1216	12.8723	15.3822	18.4957	<b>11.9180</b>
TID2013 (3000)	SRCC	0.6394	<b>0.7417</b>	0.6437	0.3977	0.6563	0.5213	<b>0.6906</b>
	PLCC	0.7017	<b>0.7895</b>	<b>0.7183</b>	0.5840	0.7175	0.6286	0.6955
	KRCC	0.4696	<b>0.5588</b>	0.4651	0.2764	0.4792	0.3658	<b>0.5152</b>
	RMSE	0.8832	<b>0.7608</b>	<b>0.8621</b>	1.0063	0.8635	0.9658	0.8907
Directed average	SRCC	0.7575	<b>0.8448</b>	0.7629	0.6402	0.7414	0.6343	<b>0.7956</b>
	PLCC	0.7870	<b>0.8672</b>	<b>0.7977</b>	0.7331	0.7720	0.6823	<b>0.7977</b>
Weighted average	SRCC	0.6881	<b>0.7842</b>	0.6928	0.4712	0.6914	0.5679	<b>0.7339</b>
	PLCC	0.7369	<b>0.8215</b>	<b>0.7510</b>	0.6455	0.7400	0.6507	0.7376

$$\text{RMSE} = \sqrt{\frac{1}{W} \sum_{i=1}^W (x_i - y_i)^2} \quad (39)$$

where  $s_i$  presents the difference between the ranks of the  $i$ -th image in subjective and objective predicted evaluations,  $\mu_x$  and  $\mu_y$  are the means of all  $x_i$  and  $y_i$ , respectively,  $W_{\text{con}}$  and  $W_{\text{dis}}$  are the numbers of concordant and discordant pairs in the database, respectively. Clearly, good performance of an IQA scheme means the values near 1 for SRCC, PLCC and KRCC, and the RMSE value close to 0. Moreover, two average measures, i.e., the directed average and the weighted average, can be both determined by the equation (40).

$$\text{AVE} = \frac{\sum_{j=1}^O v_j \omega_j}{\sum_{j=1}^O \omega_j} \quad (40)$$

where  $O$  represents the number of databases, and  $v_j$  indicates one of SRCC and PLCC values in the  $j$ -th database. As for the directed average,  $\omega_j = 1$ . As for the weighted average,  $\omega_j$  represents the number of images in the  $j$ -th database, i.e.,  $\omega_1 = 779$  and  $\omega_2 = 3000$ .

### B. RR IQA Performance Analysis

1) *Whole Performance Comparison*: In order to illustrate the advantage of the proposed hashing algorithm in RR IQA, we compare it with some popular IQA schemes, including two FR IQA schemes, i.e., PSNR [49] and SSIM [53], and four RR IQA schemes, i.e., WDWT [5], FSI [50], OSVP [54] and WNISM [55]. We summarize whole performance among different IQA schemes on the LIVE and TID2013 databases. Table VI displays the calculated IQA results, where the best performance results of FR IQA schemes and RR IQA schemes are highlighted in bold, respectively. It can be observed that the SSIM has the best results in FR IQA and our hashing algorithm has the best results in RR IQA. For the compared RR IQA schemes, our hashing algorithm achieves the biggest SRCC, PLCC, KRCC and the smallest RMSE on the LIVE database. Clearly, our hashing algorithm achieves better RR IQA performance than other RR IQA schemes on the LIVE database.

Specifically, the SRCC, PLCC, KRCC and RMSE values of the proposed hashing algorithm on the LIVE database are 0.9006, 0.8998, 0.7231 and 11.9180, respectively. The results demonstrate that the proposed hashing algorithm is superior to the compared RR IQA schemes and it is also better than the well-known FR IQA scheme called PSNR. Similarly, for the TID2013 database, the SRCC, PLCC, KRCC and RMSE values of our hashing algorithm are 0.6906, 0.6955, 0.5152 and 0.8907, respectively. It can be found that the SRCC and KRCC values of our hashing algorithm achieve the best performance among all RR IQA schemes and they are larger than those of PSNR. Note that WDWT is a useful image hashing designed for RR IQA, which exploits traditional 2D DWT to extract features from the luminance component of color image (some channels are discarded). The proposed hashing algorithm is better than WDWT on the LIVE database for all criteria. As to the TID2013 database, the PLCC and RMSE values of the proposed hashing algorithm is slightly worse than those of WDWT. But they are better than those of FSI and WNISM, illustrating good IQA performance. Therefore, the whole IQA performance of the proposed hashing algorithm is better than that of WDWT. This is due to the fact that the proposed hashing algorithm uses CCWT to exploit all information for perceptual feature extraction, but WDWT discards some channels during feature extraction. The SSIM has better IQA performance than our hashing algorithm on all databases. This is because SSIM is a FR IQA scheme which can exploit all information of the reference image to assess image quality. But our hashing algorithm is an RR IQA scheme which can only access partial information of the reference image. In some applications, full information of the reference image is unavailable. In this case, an RR IQA scheme instead of a FR IQA scheme can be used for assessing image quality.

To further examine the whole performance of the proposed hashing algorithm, the directed average and weighted average of SRCC and PLCC are illustrated in Table VI. It is obvious that the proposed hashing algorithm has bigger SRCC values in both directed average and weighted average than other RR IQA schemes. For PLCC, our weighted average is bigger

TABLE VII  
SRCC VALUES UNDER INDIVIDUAL DISTORTION TYPE ON THE LIVE AND TID2013 DATABASES

Database	Distortion type	FR IQA		RR IQA				
		PSNR	SSIM	WDWT	FSI	OSVP	WNISM	Proposed
LIVE (779)	White noise	<b>0.9854</b>	0.9694	0.9252	0.9231	0.9726	0.8701	<b>0.9726</b>
	JPEG compression	0.8809	<b>0.9764</b>	0.9284	<b>0.9623</b>	0.9297	0.9204	0.9526
	JPEG2000 compression	0.8954	<b>0.9614</b>	0.9049	0.9023	0.8967	0.9330	<b>0.9411</b>
	Gaussian blur	0.7823	<b>0.9517</b>	0.8741	<b>0.9642</b>	0.9400	0.9145	0.9141
	Fast-fading noise	0.8907	<b>0.9556</b>	0.8952	0.8863	<b>0.9460</b>	0.9227	0.8181
TID2013 (3000)	Quantization noise	<b>0.8808</b>	0.8657	0.7436	0.2618	0.6892	0.6539	<b>0.9071</b>
	Spatially correlated noise	<b>0.9200</b>	0.8515	0.8329	0.7033	0.8843	0.6726	<b>0.8982</b>
	Noise in color comp.	<b>0.8981</b>	0.7726	0.6288	0.5991	0.7589	0.6031	<b>0.8135</b>
	High frequency noise	<b>0.9140</b>	0.8634	0.8471	0.7715	0.8198	0.7469	<b>0.8785</b>
	Additive Gaussian noise	<b>0.9291</b>	0.8671	0.7855	0.7087	0.8577	0.6788	<b>0.8732</b>
	Impulse noise	<b>0.8968</b>	0.7503	0.7957	0.7045	<b>0.8281</b>	0.6771	0.8103
	JPEG2000 compression	0.8840	<b>0.9468</b>	0.9069	0.9061	0.8969	0.9182	<b>0.9239</b>
	Image denoising	<b>0.9480</b>	0.9254	0.8627	0.8312	<b>0.9162</b>	0.8584	0.9007
	Masked noise	<b>0.8323</b>	0.7767	0.3852	<b>0.7210</b>	0.4928	0.6245	0.5854
	JPEG compression	0.9189	<b>0.9200</b>	0.8869	0.8578	0.8646	0.8427	<b>0.8974</b>
	Non eccentricity pattern noise	0.6863	<b>0.7821</b>	0.6636	0.4457	0.7457	0.4483	<b>0.7995</b>
	Gaussian blur	0.9149	<b>0.9668</b>	0.8852	<b>0.9501</b>	0.9299	0.9319	0.8815
	JPEG2000 transmission errors	<b>0.8883</b>	0.8828	0.8299	0.6358	<b>0.8777</b>	0.7567	0.7424
	JPEG transmission errors	0.7685	<b>0.8493</b>	0.8275	0.3636	0.8437	<b>0.8475</b>	0.6219
	Lossy compr. of noisy images	<b>0.9145</b>	0.9057	0.8636	0.3605	0.8521	0.7208	<b>0.9150</b>
	Comfort noise	0.8411	<b>0.8566</b>	0.8248	0.5293	0.8658	0.7833	<b>0.8859</b>
	Multiplicative Gaussian noise	<b>0.8905</b>	0.7803	0.7275	0.6378	0.8246	0.5869	<b>0.8280</b>
	Intensity shift	0.7671	<b>0.7752</b>	0.4269	0.6173	0.5463	0.4910	<b>0.7300</b>
	Change of color saturation	0.0766	<b>0.4141</b>	0.0780	0.2402	0.2534	0.2774	<b>0.5319</b>
Contrast change	<b>0.4400</b>	0.3775	0.5052	<b>0.5683</b>	0.5258	0.5290	0.3917	
Local block-wise dist.	0.1552	<b>0.5720</b>	0.3484	<b>0.5584</b>	0.2946	0.2751	0.3549	
Sparse sampl. and reconstr.	0.9042	<b>0.9461</b>	0.8926	0.8815	0.9195	0.9137	<b>0.9430</b>	
Chromatic aberrations	<b>0.8872</b>	0.8775	0.8384	<b>0.8663</b>	0.8618	0.8576	0.7738	
Image color quant. w. dither	<b>0.9269</b>	0.8542	0.7952	0.2815	0.8494	0.4543	<b>0.8801</b>	
Number		15	14	0	7	4	1	17

than those of all compared RR IQA schemes, except that of WDWT, and it is also bigger than that of PSNR. In summary, our hashing algorithm achieves promising results in terms of the directed average and weighted average over the two databases.

2) *Scatter Plots*: The scatter plots of subjective scores versus objective prediction scores for different IQA schemes on the LIVE database are displayed in Figure 11, where a blue “+” indicates one test image, and red curves are the fitted curves estimated by the logistic regression analysis [56]. The LIVE database provides subjective scores (i.e., DMOS) and the IQA schemes generate objective prediction scores. The representative IQA schemes include PSNR [49], WDWT [5], FSI [50], OSVP [54], WNISM [55] and the proposed hashing algorithm. As can be seen from the Figure 11, the points of the proposed hashing algorithm concentrate around the fitted curve, indicating that the objective prediction scores of our hashing algorithm have better consistency with subjective scores than those of the compared IQA schemes.

3) *Individual Distortion Comparison*: To further comprehensively verify RR IQA performance, the predicted ability of the proposed hashing algorithm for individual distortion types

is demonstrated. Here, we give the SRCC among all evaluation criteria. This is because the SRCC indicates the convergency and monotonicity between subjective and objective predicted evaluations. Table VII lists SRCC values under individual distortion types on two databases. For simplicity, the best results in FR IQA and RR IQA schemes are highlighted in bold, respectively. It can be found that SSIM and PSNR in FR IQA scheme are marked 14 and 15 times, respectively. The proposed algorithm is marked 17 times, ranked first in RR IQA schemes. More concretely, there are 11 kinds of noise distortions on the two databases. Compared with other RR IQA schemes, our hashing algorithm reaches the best results in these noise distortions, except for three kinds of noises, i.e., fast-fading noise, impulse noise and masked noise. Furthermore, our hashing algorithm outperforms the classical FR-IQA schemes, i.e., PSNR and SSIM, in many distortion types, such as lossy compression of noisy images, quantization noise, comfort noise and change of color saturation. Consequently, it is further confirmed that the proposed hashing algorithm reaches the advanced performance for individual distortion types. The proposed hashing algorithm achieves a better IQA performance of individual distortions than the

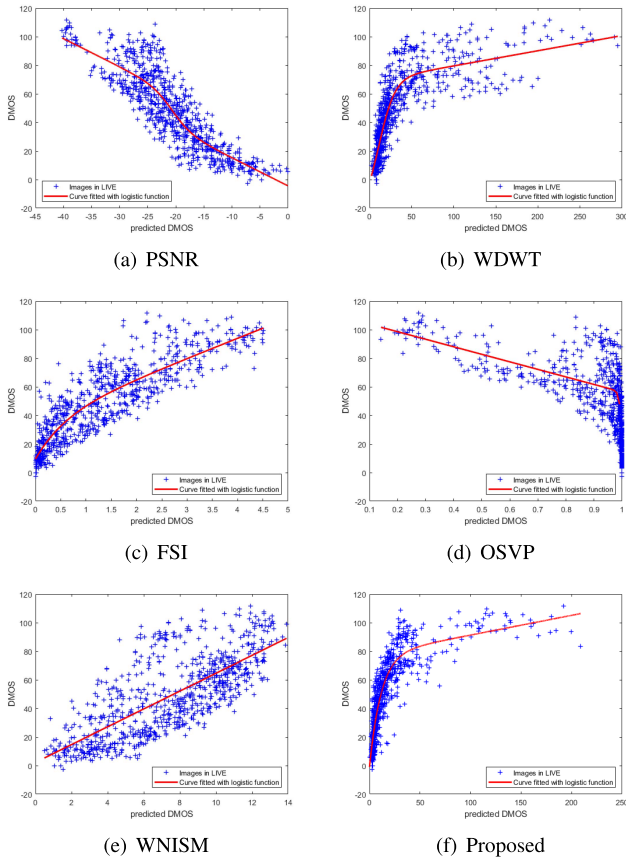


Fig. 11. Comparison of scatter plots on the LIVE database.

compared schemes. This can be understood as follows. The calculation of CCWT uses all color channels without discarding any information and thus the distortions on color channels are well preserved in the CCWT sub-bands. Therefore, our color features extracted from these sub-bands can effectively describe image distortion on color channels.

## VI. CONCLUSION

An effective image hashing algorithm with CCWT and CS has been presented in this paper, which reaches good performances of classification and RR IQA application. A key contribution is the use of CCWT to decompose input color image into different sub-bands. Since the calculation of CCWT uses all color channels without discarding any information, the distortions of color channels are preserved in the CCWT sub-bands. Another major contribution is the perceptual feature extraction from the CCWT sub-bands by the block-based CS. As the Euclidean distance between the block-based CS features is slightly influenced by content-preserving operations, feature construction by Euclidean distances can guarantee robust, discriminative and compact. Numerous experiments have been done for verifying the efficiency of our hashing algorithm. The results have demonstrated that our hashing algorithm can achieve a desirable classification performance. Besides, the results of experiments on two open databases have validated the superiority of our hashing algorithm in RR IQA application.

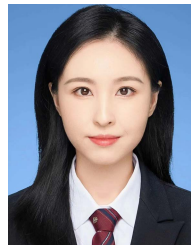
## ACKNOWLEDGMENT

The authors would like to thank to the reviewers for their comments and suggestions.

## REFERENCES

- [1] Z. Huang and S. Liu, "Perceptual hashing with visual content understanding for reduced-reference screen content image quality assessment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 7, pp. 2808–2823, Jul. 2021.
- [2] 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.
- [3] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *J. Visual Commun. Image Representation*, vol. 22, no. 4, pp. 297–312, 2011.
- [4] X. Lv and Z. J. Wang, "Reduced-reference image quality assessment based on perceptual image hashing," in *Proc. 16th IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 4301–4304.
- [5] Z. Tang, Z. Huang, H. Yao, X. Zhang, L. Chen, and C. Yu, "Perceptual image hashing with weighted DWT features for reduced-reference image quality assessment," *Comput. J.*, vol. 61, no. 11, pp. 1695–1709, Nov. 2018.
- [6] C. Qin, E. Liu, G. Feng, and X. Zhang, "Perceptual image hashing for content authentication based on convolutional neural network with multiple constraints," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 11, pp. 4523–4537, Nov. 2021.
- [7] J. Zhang and Y. Peng, "SSDH: Semi-supervised deep hashing for large scale image retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 1, pp. 212–225, Jan. 2019.
- [8] Z. Tang, L. Chen, X. Zhang, and S. Zhang, "Robust image hashing with tensor decomposition," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 3, pp. 549–560, Mar. 2019.
- [9] X. Lu, X. Zheng, and X. Li, "Latent semantic minimal hashing for image retrieval," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 355–368, Jan. 2017.
- [10] C.-P. Yan, C.-M. Pun, and X.-C. Yuan, "Quaternion-based image hashing for adaptive tampering localization," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 12, pp. 2664–2677, Dec. 2016.
- [11] R. Venkatesan, S.-M. Koon, M. H. Jakubowski, and P. Moulin, "Robust image hashing," in *Proc. Int. Conf. Image Process.*, 2000, pp. 664–666.
- [12] X. Wang, N. Zheng, J. Xue, and Z. Liu, "A novel image signature method for content authentication," *Comput. J.*, vol. 55, no. 6, pp. 686–701, Jun. 2012.
- [13] F. Ahmed, M. Y. Siyal, and V. U. Abbas, "A secure and robust hash-based scheme for image authentication," *Signal Process.*, vol. 90, no. 5, pp. 1456–1470, 2010.
- [14] C. Qin, C.-C. Chang, and P.-L. Tsou, "Robust image hashing using non-uniform sampling in discrete Fourier domain," *Digit. Signal Process.*, vol. 23, no. 2, pp. 578–585, 2013.
- [15] A. Swaminathan, Y. Mao, and M. Wu, "Robust and secure image hashing," *IEEE Trans. Inf. Forensics Security*, vol. 1, no. 2, pp. 215–230, Jun. 2006.
- [16] C. Plesca and L. Morogan, "Efficient and robust perceptual hashing using log-polar image representation," in *Proc. 10th Int. Conf. Commun. (COMM)*, May 2014, pp. 1–6.
- [17] Y. Liu, G. Xin, and Y. Xiao, "Robust image hashing using radon transform and invariant features," *Radioengineering*, vol. 25, no. 3, pp. 556–564, Sep. 2016.
- [18] V. Monga and B. L. Evans, "Perceptual image hashing via feature points: Performance evaluation and trade-offs," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3453–3466, Oct. 2006.
- [19] L. Wang, X. Jiang, S. Lian, D. Hu, and D. Ye, "Image authentication based on perceptual hash using Gabor filters," *Soft Comput.*, vol. 15, no. 3, pp. 493–504, Mar. 2011.
- [20] Y. Zhao, S. Wang, X. Zhang, and H. Yao, "Robust hashing for image authentication using Zernike moments and local features," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 1, pp. 55–63, Jan. 2013.
- [21] X. Wang, K. Pang, X. Zhou, Y. Zhou, L. Li, and J. Xue, "A visual model-based perceptual image hash for content authentication," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 7, pp. 1336–1349, Jul. 2015.

- [22] Z. Tang *et al.*, “Robust image hashing via visual attention model and ring partition,” *Math. Biosci. Eng.*, vol. 16, no. 5, pp. 6103–6120, 2019.
- [23] S. S. Kozat, R. Venkatesan, and M. K. Mihcak, “Robust perceptual image hashing via matrix invariants,” in *Proc. Int. Conf. Image Process. (ICIP)*, 2004, pp. 3443–3446.
- [24] V. Monga and M. K. Mihcak, “Robust image hashing via non-negative matrix factorizations,” in *Proc. IEEE Int. Conf. Acoust. Speed Signal Process.*, May 2006, pp. 225–228.
- [25] L.-W. Kang, C.-S. Lu, and C.-Y. Hsu, “Compressive sensing-based image hashing,” in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 1285–1288.
- [26] Z. Tang, X. Zhang, X. Li, and S. Zhang, “Robust image hashing with ring partition and invariant vector distance,” *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 1, pp. 200–214, Jan. 2016.
- [27] C. Qin, M. Sun, and C.-C. Chang, “Perceptual hashing for color images based on hybrid extraction of structural features,” *Signal Process.*, vol. 142, pp. 194–205, Jan. 2018.
- [28] S. P. Singh, G. Bhatnagar, and A. K. Singh, “A new robust reference image hashing system,” *IEEE Trans. Dependable Secure Comput.*, vol. 19, no. 4, pp. 2211–2225, Jul. 2022.
- [29] Y. Li, Z. Lu, C. Zhu, and X. Niu, “Robust image hashing based on random Gabor filtering and dithered lattice vector quantization,” *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1963–1980, Apr. 2012.
- [30] X. Huang, X. Liu, G. Wang, and M. Su, “A robust image hashing with enhanced randomness by using random walk on zigzag blocking,” in *Proc. IEEE Trustcom/BigDataSE/ISPA*, Aug. 2016, pp. 14–18.
- [31] Y. Li and L. Guo, “Robust image fingerprinting via distortion-resistant sparse coding,” *IEEE Signal Process. Lett.*, vol. 25, no. 1, pp. 140–144, Jan. 2018.
- [32] C. Zheng, L. Zhu, X. Lu, J. Li, Z. Cheng, and H. Zhang, “Fast discrete collaborative multi-modal hashing for large-scale multimedia retrieval,” *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 11, pp. 2171–2184, Nov. 2020.
- [33] B. Yang and S. Li, “Multifocus image fusion and restoration with sparse representation,” *IEEE Trans. Instrum. Meas.*, vol. 59, no. 4, pp. 884–892, Apr. 2010.
- [34] B. Ahmaderaghi, F. Kurugollu, and J. M. Del Rincon, “Blind image watermark detection algorithm based on discrete shearlet transform using statistical decision theory,” *IEEE Trans. Comput. Imag.*, vol. 4, no. 1, pp. 46–59, Mar. 2018.
- [35] Y. Chen, D. Li, and J. Q. Zhang, “Complementary color wavelet: A novel tool for the color image/video analysis and processing,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 1, pp. 12–27, Jan. 2019.
- [36] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [37] A. Griffin and P. Tsakalides, “Compressed sensing of audio signals using multiple sensors,” in *Proc. 16th Eur. Signal Process. Conf.*, Aug. 2008, pp. 1–5.
- [38] D. J. Brady, K. Choi, D. L. Marks, R. Horisaki, and S. Lim, “Compressive holography,” *Opt. Exp.*, vol. 17, no. 15, pp. 13040–13049, 2009.
- [39] M. Lustig, D. Donoho, and J. M. Pauly, “Sparse MRI: The application of compressed sensing for rapid MR imaging,” *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, Oct. 2007.
- [40] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [41] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, “Contour detection and hierarchical image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.
- [42] Y. Gan and Q. Zhao, “An effective defect inspection method for LCD using active contour model,” *IEEE Trans. Instrum. Meas.*, vol. 62, no. 9, pp. 2438–2445, Sep. 2013.
- [43] Z. Tang, L. Huang, X. Zhang, and H. Lao, “Robust image hashing based on color vector angle and Canny operator,” *AEU Int. J. Electron. Commun.*, vol. 70, no. 6, pp. 833–841, 2016.
- [44] L. Li, S. Wang, W. Zeng, and F. Qin, “Research on zero watermarking technique based on improved Sobel operator,” in *Proc. IEEE Int. Conf. Artif. Intell. Ind. Design (AIID)*, May 2021, pp. 594–597.
- [45] G. Schaefer and M. Stich, “UCID: An uncompressed color image database,” in *Proc. SPIE*, vol. 5307, pp. 472–480, Dec. 2004.
- [46] F. A. P. Petitcolas, “Watermarking schemes evaluation,” *IEEE Signal Process. Mag.*, vol. 17, no. 5, pp. 58–64, Sep. 2000.
- [47] T.-Y. Lin *et al.*, “Microsoft COCO: Common objects in context,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 740–755.
- [48] T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.
- [49] A. Tanchenko, “Visual-PSNR measure of image quality,” *J. Vis. Commun. Image Represent.*, vol. 25, no. 5, pp. 874–878, 2014.
- [50] Y. Liu, G. Zhai, K. Gu, X. Liu, D. Zhao, and W. Gao, “Reduced-reference image quality assessment in free-energy principle and sparse representation,” *IEEE Trans. Multimedia*, vol. 20, no. 2, pp. 379–391, Feb. 2018.
- [51] H. R. Sheikh, K. Seshadrinathan, A. K. Moorthy, Z. Wang, A. C. Bovik, and L. K. Cormack. (2006). *Image and Video Quality Assessment Research at Live*. [Online]. Available: <http://live.ece.utexas.edu/research/quality/>
- [52] N. Ponomarenko *et al.*, “Color image database TID2013: Peculiarities and preliminary results,” in *Proc. Eur. Workshop Vis. Inf. Process. (EUVIP)*, Jun. 2013, pp. 106–111.
- [53] 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.
- [54] J. Wu, W. Lin, G. Shi, L. Li, and Y. Fang, “Orientation selectivity based visual pattern for reduced-reference image quality assessment,” *Inf. Sci.*, vol. 351, pp. 18–29, Jul. 2016.
- [55] Z. Wang and E. P. Simoncelli, “Reduced-reference image quality assessment using a wavelet-domain natural image statistic model,” *Proc. SPIE*, vol. 5666, pp. 149–159, Mar. 2005.
- [56] J. Kim *et al.*, “Optimal feature selection for pedestrian detection based on logistic regression analysis,” in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2013, pp. 239–242.



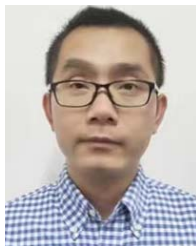
**Mengzhu Yu** received the B.S. degree from the Qingdao University of Science and Technology, Qingdao, China, in 2017, and the M.Eng. degree from Guangxi Normal University, Guilin, China, in 2020, where she is currently pursuing the Ph.D. degree. Her research interests include image processing and multimedia security.



**Zhenjun Tang** (Member, IEEE) received the B.S. and M.Eng. degrees from Guangxi Normal University, Guilin, China, in 2003 and 2006, respectively, and the Ph.D. degree from Shanghai University, Shanghai, China, in 2010. He is a Professor with the School of Computer Science and Engineering, Guangxi Normal University. He has contributed more than 90 international journal papers. He is a reviewer of more than 30 SCI journals, such as IEEE TRANSACTIONS, Elsevier journals, and Springer journals. His research interests include image processing, video processing, and multimedia security.



**Xianquan Zhang** received the M.Eng. degree from Chongqing University, Chongqing, China, in 1996. He is currently a Professor with the School of Computer Science and Engineering, Guangxi Normal University. He has contributed more than 100 articles. His research interests include image processing and multimedia security.



**Bineng Zhong** received the B.S., M.S., and Ph.D. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 2004, 2006, and 2010, respectively. From 2007 to 2008, he was a Research Fellow at the Institute of Automation and the Institute of Computing Technology, Chinese Academy of Science. From September 2017 to September 2018, he was a Visiting Scholar at Northeastern University, Boston, MA, USA. From November 2010 to October 2020, he was a Professor at the School of Computer Science and Technology,

Huaqiao University, Xiamen, China. Currently, he is a Professor with the School of Computer Science and Engineering, Guangxi Normal University. His current research interests include pattern recognition, machine learning, and computer vision.



**Xinpeng Zhang** (Member, IEEE) received the B.S. degree in computational mathematics from Jilin University, China, in 1995, and the M.E. and Ph.D. degrees in communication and information system from Shanghai University, China, in 2001 and 2004, respectively. He was a Visiting Scholar at the State University of New York, Binghamton, from 2010 to 2011, and also with Konstanz University as an experienced Researcher, sponsored by the Alexander von Humboldt Foundation, from 2011 to 2012. Since 2004, he has been a Faculty Member with the School of Communication and Information Engineering, Shanghai University, where he is currently a Professor. He is also a Faculty Member with the School of Computer Science, Fudan University. He has published over 200 papers in his research areas. His research interests include multimedia security, image processing, and digital forensics. From 2014 to 2017, he was an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY.