Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8820436, 14 pages
https://doi.org/10.1155/2020/8820436

*Research Article*

# A Novel DIBR 3D Image Hashing Scheme Based on Pixel Grouping and NMF

**Chen Cui**,[1,2] **Xujun Wu**,[1] **Jun Yang**,[3] **and Juyan Li**[1,4]

[1]*School of Information Science and Technology, Heilongjiang University, Harbin 150080, China*
[2]*Guangxi Key Laboratory of Cryptography and Information Security, Guilin 541004, China*
[3]*College of Mathematics Physics and Information Engineering, Jiaxing University, Jiaxing 314000, China*
[4]*State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China*

Correspondence should be addressed to Juyan Li; lijuyan587@163.com

Received 25 June 2020; Revised 4 November 2020; Accepted 25 November 2020; Published 10 December 2020

Academic Editor: Ding Wang



Most of the traditional 2D image hashing schemes do not take into account the change of viewpoint when constructing the final hash vector. This result in the classification accuracy rate is unsatisfactory when applied for depth-image-based rendering (DBIR) 3D image identification. In this work, pixel grouping based on histogram shape and nonnegative matrix factorization (NMF) are applied to design DIBR 3D image hashing with better robustness resisting to geometric distortions and higher classification accuracy rate for virtual image identification. Experiments show that the proposed hashing is robust against common signal and geometric distortion attacks, such as additive noise, blurring, JPEG compression, scaling, and rotation. Compared with the state-of-art schemes of traditional 2D image hashing, the proposed hashing achieves better performances under above attacks, especially for virtual image identification.

## 1. Introduction

Depth-image-based rendering (DIBR) [1] is a kind of 3D representation technology, by which the virtual left and right images are generated from the center image according to the depth information described with the depth image. Then, viewers can easily get stereo perception with the virtual image pair. In the digital communication model of DIBR 3D image, receiver performs depth-image-based rendering operation to generate virtual image pair for 3D video perception. As a matter of fact, either of the center image, the virtual left image and the virtual right image may suffer from illegal or unauthorized redistribution. In order to resolve this problem, robust perceptual hashing has been widely used for digital multimedia protection. As variety of copies for center image and virtual images existing, image hashing can also help us to find the similar one and detect the tempered [2–6]. In this paper, we focus on designing a robust image hashing scheme for DIBR 3D image identification.

In the DIBR system, virtual right image and left image are generated from the corresponding center image with pixel mapping. In a sense, virtual images have similar visual content with their corresponding center image, which demands the hashing scheme should identify the virtual images with the same content as the center image as shown in Figure 1.

Generally, traditional 2D image hashing should have the several characteristics such as one-way function, compactness, perceptual robustness, visual fragility, and unpredictability [7]. For DIBR 3D image hashing, the perceptual robustness should have more stringent requirements as

$$\begin{aligned} P(H_k(I_c) \approx H_k(I_v)) &\geq 1-\varepsilon, 0 \leq \varepsilon < 1, \\ P(H_k(I_c) \approx H_k(I_d)) &\geq 1-\tau, 0 \leq \tau < 1. \end{aligned} \tag{1}$$

$I_c$ represents the center image, $I_v$ represents the virtual image, and $I_d$ represents the perceptually similar copy of $I_c$ or $I_v$ with minor distortion. Here, $\varepsilon$ and $\tau$ should be close

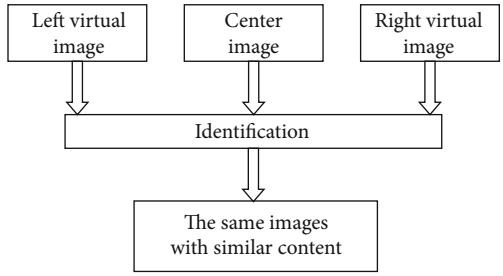


Figure 1: The character of image hashing for DIBR 3D images.

to zero. This paper focuses on designing a robust DIBR 3D image hashing for DIBR 3D image identification.

## 2. Related Work

Image hashing, a technique to derive a security content-based compact signature for input image [8], has been extensively studied for identification [9, 10], authentication [11–14], quality assessment [15], and tampering detection [5, 16–18].

In general, robustness and discrimination are two important aspects should be considered to design an image hashing scheme. Robust image hashing has been extensively studied for content-based identification for traditional 2D images. As feature extraction affects the identification performance for image hashing, many existing methods focus on extracting robust features resisting to content-preserving operations [9, 19]. In addition, some dimensionality reduction methods have been adopted to extract the robust features for hash generation, such as singular value decomposition (SVD) [20] and nonnegative matrix factorization (NMF) [21], which are robust to most kinds of signal distortion attacks but sensitive to geometric distortions such as rotation. In [22], a robust image hashing with multidimensional scaling is proposed, which achieves better performance when taking into account image classification. In order to make the hashing scheme robust to geometric distortion attacks such as rotation, robust image hashing scheme is proposed to extract the geometric-invariant features for generating the final hash vector [23]. In [24], a robust image hash in Radon transform domain is proposed, which is robust against rotation, but the discriminative capability is not good enough. In [25], invariant moments extracted from color spaces are used to generate the final hash vector. This hashing scheme is robust against rotation, but increase misclassification. In [26], Li uses Gabor filtering to extract features and compresses these features with dithered lattice vector quantization to generate the compact hash. This method is robust against rotation, but the discriminative capability is also not good enough.

In recent years, some novel and excellent hashing algorithms are proposed. Qin et al. [27] propose a security image hashing scheme based on perceptual texture and structure features, but the image classification performance is not good enough. In [28], a robust image hashing based on tensor decomposition is proposed, which is robust to common signal distortion attacks. However, the discriminative capability is not good enough. Lv et al. [7] propose shape contexts and local feature point-based image hashing scheme. Compressing the descriptors of SIFT feature points in each hash bin to form the final hash vector, their hashing scheme is robust to geometric distortion attacks such as rotation. However, the performance is degraded when the detected key points from the test image are not stable enough to coincide with the detected key points from the original. Tang et al. [29, 30] propose a kind of robust image hashing scheme based on ring partition. Using the pixels in each ring to form a secondary image insensitive to rotation, they extract the final hash vector from the secondary image. The experimental results show that their hash schemes are robust to rotation with good discriminative capability. This kind of method considers that the viewpoint never changes when the digital image is attacked by most of the content-preserving manipulations. In other words, the image center of original image and their copies would not change.

Performance comparisons among some traditional 2D image hashing algorithms are summarized in Table 1. For signal distortion attacks, the word "Yes" means that the algorithm is robust against some operations including additive noise, blurring, and JPEG compression. For geometric distortion attacks, the word "Yes" means that the algorithm is robust against scaling, rotation within arbitrary degree, and the word "Unknown" means that such performance result has not been reported in the literature as far as we know.

In fact, the image center of center image and virtual images are different, which is caused by the DIBR operations. As a result, this kind of traditional 2D hashing scheme would not achieve good performance when applied for DIBR 3D image identification. Some of the state-of-art traditional 2D robust image hashing schemes resisting to geometric distortions do not take into account the situation about viewpoint changing [7, 29, 30]. Dividing the image into several rings or constructing rotation-invariant secondary image according to the unchanged image center is the key step to construct hash vector robust to rotation manipulation. However, the image center changes when generating virtual images in the DIBR system.

In this work, a pixel grouping and nonnegative matrix factorization-based hashing scheme is designed for DIBR 3D image identification. The key contribution is using the approximate invariance of histogram shape to extract features insensitive to the operation of virtual image generation, making our DIBR 3D image hashing scheme identify the virtual images with the same visual content as the original center image. The rest of this paper is organized as follows: Section 2 briefly reviews the DIBR operations. Section 3 introduces the pixel grouping according to approximate invariance of histogram shape and nonnegative matrix factorization-based image hashing. Section 4 shows the experimental results and performance comparisons. Section 5 gives the final conclusions.

## 3. Review of Depth-Image-Based Rendering Process

Figure 2 illustrates the relationship between the center image and the virtual images generated by DIBR operations [31]. Suppose $P$ is a point in the space, $C_c$, $C_l$, and $C_r$ represent

Table 1: Performance comparisons among some typical algorithms.

| Algorithm | Against common content-preserving operations | | | | | Discriminative capability |
|---|---|---|---|---|---|---|
| | JPEG compression | Additive noise | Blurring | Image rotation | Image scaling | |
| [7] | Yes | Yes | Yes | Yes | Yes | Unknown |
| [8] | Yes | Unknown | Yes | Unknown | Yes | Good |
| [11] | Yes | Unknown | Unknown | No | No | Unknown |
| [17] | Yes | Yes | Yes | Yes | Yes | Unknown |
| [18] | Yes | Unknown | Yes | Yes | Yes | Moderate |
| [20] | Yes | Yes | Yes | No | Yes | Poor |
| [21] | Yes | Yes | Yes | No | Yes | Poor |
| [24] | Yes | Yes | Yes | Yes | Yes | Moderate |
| [25] | Yes | Yes | Yes | Yes | Yes | Poor |
| [26] | Yes | Yes | Yes | No | Yes | Good |
| [27] | Yes | Unknown | Yes | No | Yes | Moderate |
| [28] | Yes | Yes | Yes | No | Yes | Moderate |
| [29] | Yes | Yes | Yes | Yes | Yes | Good |
| [30] | Yes | Yes | Yes | Yes | Yes | Good |

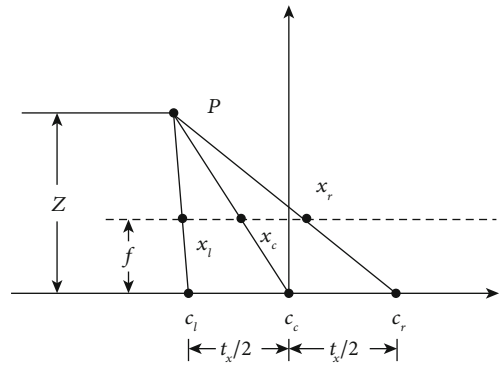


Figure 2: The relationship of pixel in the left image, center image, and right image.

the center viewpoint, left viewpoint, and the right viewpoint, respectively, $f$ represents the focal length of the center viewpoint, and $Z$ represents the depth of $P$. $x_c$, $x_l$, and $x_r$ represent the $x$-coordinate of pixel in the center image, the virtual left image, and the virtual right image, respectively. $t_x$ represents the baseline distance, value of which is equal to the distance between the left and right viewpoints. As geometric relations shown in Figure 2, $x$-coordinate of pixel in the virtual images is computed as

$$x_l = x_c + \frac{t_x}{2}\frac{f}{Z}, \qquad x_r = x_c - \frac{t_x}{2}\frac{f}{Z}, \tag{2}$$

$$Z(v) = Z_{\text{far}} + v \times \frac{Z_{\text{near}} - Z_{\text{far}}}{255}, \; v \in [0, 255]. \tag{3}$$

In fact, the gray value of pixel in depth image is not the real depth value. Pixel with gray value close to 255 indicates that $P$ is close to the near clipping plane $Z_{\text{near}}$. On the other hand, pixel with gray value close to 0 indicates that $P$ is close to the far clipping plane $Z_{\text{far}}$. According to formula (3), the depth value $Z(v)$ of $P$ is computed, where $v$ represents the gray value.

## 4. Proposed Image Hashing

Our DIBR 3D image hashing scheme includes the following steps: the original center image is filtered with a Gaussian kernel low-pass filter to get the low frequency, and we standardize the low frequency of center image for hash generation. Then, pixels of normalized low frequency image are divided into different groups according to the histogram shape. Then, these pixel groups are used to construct a secondary image, which is almost unchangeable under geometric distortions and slightly changes after DIBR operations. Lastly, the secondary image is discomposed by nonnegative matrix factorization to get the coefficient matrix, and the final hash is constructed with these coefficients.

*4.1. Preprocessing.* Low-pass filtering is adopted to extract the low-frequency component of original center image, which is aimed at enhancing the robustness of proposed hashing

scheme to some common content-preserving manipulations [32]. The low-frequency component $IC_{low}$ of original center image IC is obtained as

$$IC_{low}(x, y) = G(x, y, \sigma) * I(x, y), \quad (4)$$

where $*$ represents the convolution operation, and the low-pass filter Gaussian function $G(x, y, \sigma)$ is represented as

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-x^2+y^2/2\sigma^2}, \quad (5)$$

where $\sigma$ is the standard difference. According to parameters setting in [32], $\sigma$ is set to 1.

*4.2. Pixel Grouping.* The gray levels of filtered image $I_{low}$ also range from 0 to 255. In this paper, only pixels with $M$ different gray levels are randomly selected to construct the secondary image, which is aimed at ensuring the security of proposed hashing algorithm. With a key-based sequence $P(M) = \{p_i | i = 1 \cdots M, 0 \le p_i \le 255\}$, $M$ gray levels $h_1, h_2, \cdots, h_M$ are selected for pixels grouping, where $h_i = p_i$. The set of selected gray level is represented as

$$H_M = \{h_i | i = 1, 2, 3, \cdots \cdots, M\}. \quad (6)$$

After resizing $I_{low}$ to $m \times m$, pixels with $L_B$ neighbouring gray levels in $H_M$ are selected to form one pixel group. In total, $n = \lfloor M/L_B \rfloor$ groups are formed, where $\lfloor \cdot \rfloor$ is a floor function.

Suppose $g_i$ be one of the pixel groups. In order to form the $i^{th}$ column of the secondary image, we sort and resize $g_i$ to a new vector $v_i$ sized $k \times 1$. Then, the secondary image is represented as

$$V = [v_1, v_2, v_3, \cdots, v_n]. \quad (7)$$

It is clear that the histogram shape of $V$ is the same as that of the resized $I_{low}$, and the secondary image $V$ is robust to geometric distortions such as rotation. In this paper, $M$ is set to 240, $m = 256$, $L_B = 6$, and $k = 4m$.

*4.3. Hash Generation.* Since the histogram shape is almost unchangeable under geometric distortions and slightly changes after DIBR operations, features extracted from the secondary image $V$ also have this property. NMF is used to get the base matrix $W$ and coefficient matrix $H$, respectively. Concatenate the coefficient matrix $H$ to obtain the final hash vector, the length $L$ of hash vector is $n \times r$, where $n$ is the number of pixel groups and $r$ is the rank for NMF. In this paper, $r$ is set to 2.

In this paper, correlation coefficient is taken as the metric to measure the similarity between two image hash vectors Hash1 and Hash2. The correlation coefficient $S(\text{Hash1}, \text{Hash2})$ is defined as

$$S(\text{Hash1}, \text{Hash2}) = \frac{\text{cov}(\text{Hash1}, \text{Has}h2)}{\sqrt{D(\text{Hash1})}\sqrt{D(\text{Hash2})}}. \quad (8)$$

Table 2: Perceptual distance between center image and left virtual image computed by different hashing methods.

| Image | Proposed method | Method in [30] |
|---|---|---|
| Breakdancers | 0.9984 | -0.3817 |
| Dolls | 0.9952 | -0.7150 |
| Books | 0.9982 | -0.7499 |
| Ballet | 0.9984 | 0.8183 |

Table 3: Perceptual distance between center image and right virtual image computed by different hashing methods.

| Image | Proposed method | Method in [30] |
|---|---|---|
| Breakdancers | 0.9990 | 0.5647 |
| Dolls | 0.9909 | 0.7812 |
| Books | 0.9982 | 0.8849 |
| Ballet | 0.9980 | 0.9093 |

According to formula (8), $S(\text{Hash1}, \text{Hash2})$ ranges from $-1$ to 1, and a bigger $S(\text{Hash1}, \text{Hash2})$ value indicates that the input image is more similar with the original corresponding center image. If the correlation coefficient $S(\text{Hash1}, \text{Hash2})$ is higher than the threshold predefined, the input image is viewed as perceptual content unchanged. If the correlation coefficient $S(\text{Hash1}, \text{Hash2})$ is lower than the threshold predefined, the input image is viewed as a different image or a maliciously tempered version of the original corresponding center image. For DIBR 3D images, the virtual images should have much bigger $S(\text{Hash1}, \text{Hash2})$ value when computing the perceptual distance from their corresponding original center image. According to experiment results listed in Tables 2 and 3, some virtual images are viewed different from the original center image when the hashing method proposed in [30] is adopted. It is clear that our DIBR 3D image hashing scheme can identify the virtual images with the same visual content as the original center image.

*4.4. Invariance of Histogram Shape.* Robustness to geometric distortion attacks, especially the rotation attacks, is the major problem to be considered when designing a traditional 2D image hashing scheme with features insensitive to geometric operations. According to [33], the histogram shape is robust to scaling, rotation, and affine attacks. To design a DIBR 3D image hashing scheme, the robustness to the operation of virtual image generation is also important.

The resistance of the histogram shape to the operation of virtual image generation is discussed as follows. According to [33], with some regions cropped from the original image, the histogram shape of the original image will be different from the histogram shape of cropped one. Strictly speaking, the robustness of histogram shape under cropping attacks depends on the image and the cropped area. So the invariance property of the histogram shape of an image under cropping attacks is an approximate invariance. Similarly, in the operation of virtual image generation, the virtual images are generated from the center image with some regions cropped and

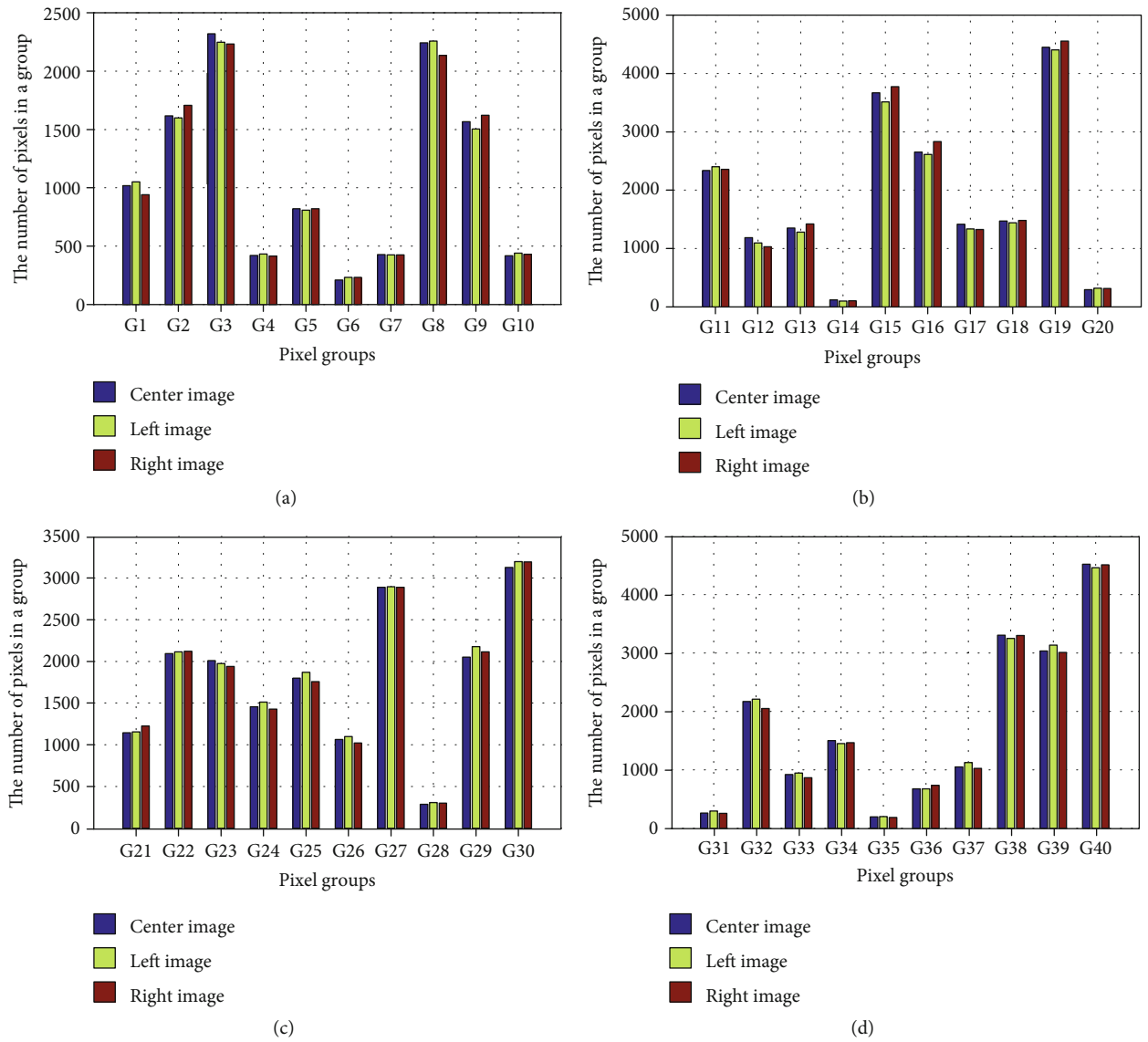


Figure 3: The number of pixels in different groups of center image and virtual images.

holes filled, and the robustness of histogram shape under this operation depends on the image, the baseline distance, and the key-based sequence used for selecting the gray levels, so the invariance property of the histogram shape of the virtual images is also an approximate invariance. As shown in Figure 3, although the virtual images are generated from the center image with pixels' translation and parts of pixels cropped, the number of pixels in each group slightly changes compared with that of the center image. Resizing each pixel group to a new vector as the column of secondary image, this secondary image is similar with that formed from the center image. Using the NMF to extract features from the secondary image to obtain the final hash vector, the final hash vector of the virtual image is almost the same as that of the center image. Table 4 illustrates the statistics of perceptual distances between the tested center images and their corresponding virtual images. It can be seen that all means are close to 1, and their standard deviations are small. Moreover, the minimum values are also close to 1. This indicates that the approximate invariance of histogram shape can be used to extract features insensitive to DIBR operations, making our DIBR 3D image hashing scheme identify the virtual images with the same visual content as the original center image.

Table 4: Statistics of perceptual distance between the center image and the virtual image.

| | Max | Min | Mean | Standard deviation |
|---|---|---|---|---|
| Left image | 0.9994 | 0.9938 | 0.9970 | 0.0022 |
| Right image | 0.9990 | 0.9893 | 0.9961 | 0.0036 |

## 5. Experimental Results

Dataset with 2727 images is constructed to evaluate the identification performance for DIBR 3D image. Pairs of the center

images and their corresponding depth images are selected from Middlebury Stereo Datasets [34] and Microsoft Research 3D Video Datasets [35] to construct the dataset, and the sizes of these images are ranging from $450 \times 375$ to $1390 \times 1110$. Hashes of the center image, the virtual left image, the virtual right image, and their distorted versions are generated with our hashing scheme in order to calculate the identification accuracy rate. The distorted versions are generated by attacking the center and virtual images according to 10 classes of common content-preserving operations. In this paper, MATLAB is exploited to implement these 10 class operations with different parameters. These operations include common signal and geometric distortion attacks such as JPEG compression, blurring, additive noise, scaling, rotation, and cropping after rotation. The operations and their parameters are listed in Table 5.

*5.1. Discrimination.* 120 different color images are collected from the Ground Truth Database [36] in order to test the discriminative capability of proposed hashing. The hash vectors are generated for these 120 images, and then 7140 correlation coefficients of $S$ are computed between each pair of different hash vectors. The maximum value of these correlation coefficients is 0.9785, and the minimum value is -0.5101. If the threshold $T$ is set as 0.92, 0.32 percent pairs of different images are identified with the similar content. 0.09 percent pairs of different images are identified with the similar content with $T$ is set to 0.94. No pair of different images is identified with the similar content when $T$ is set to 0.98.

*5.2. Perceptual Robustness.* Firstly, four pairs of the center image and the depth image are selected from the above dataset. They are "Breakdancers," "Books," "Dolls," and "ballet" as listed in Table 2. Each virtual image pair and the center image are attacked by the content-preserving operations listed in Table 5. As shown in Figure 4, no pair of visually identical images (including the distorted center and virtual images) is identified with different content when the threshold $T$ is set to 0.96.

In this paper, combinational attacks between image geometric distortion attacks and signal distortion attacks are also performed for many images to evaluate the perceptual robustness of the proposed image hashing scheme. Combinational attacks are used as follows: Gaussian noise +rotation, Gaussian noise+cropping after rotation, salt and paper noise+rotation, salt and paper noise+cropping after rotation, speckle noise+rotation, speckle noise+cropping after rotation, Gaussian blurring+rotation, Gaussian blurring+cropping after rotation, circular blurring+rotation, circular blurring+cropping after rotation, motion blurring +rotation, motion blurring+cropping after rotation, JPEG compression+rotation, and JPEG compression+cropping after rotation. In addition, scaling+rotation and scaling +cropping after rotation are also performed.

To obtain the versions under combinational attacks, both of the center image and the virtual image are firstly attacked by rotation (2°, 10°, and 45°) or cropping after rotation (2°, 10°, and 45°). To simulate combinational attacks, all operations listed in Table 5 (the quality factor of JPEG compression is set from 30 to 100) are performed except rotation and cropping after rotation. Then, hash vectors of the attacked versions are generated to compute the perceptual distances represented by the correlation coefficient $S$. For space limitation, a typical example is exemplified here. Figures 5 and 6 illustrate the robustness of our hashing against combinational attacks, where the $x$-axis is the parameter value of each manipulation, the $y$-axis is the correlation coefficient $S$, and the center image and the virtual image are firstly attacked with rotation 45° or cropping after rotation 45°, then further attacked by all operations listed in Table 5, except rotation and cropping after rotation. It is observed that all correlation coefficients are above 0.94, except the combinational attack Gaussian noise with variance 0.005 + rotation with 45°. This means that our hashing is also robust against most of the above combinational attacks. As shown in Table 6, the correlation coefficients are above 0.98, when the angle of rotation is 2°. The experiments demonstrate that our DIBR 3D hashing is robust against these combinational attacks.

Table 5: Content-preserving operations and the parameters setting.

| Manipulation | Parameter setting | Copies |
|---|---|---|
| | Additive noise | |
| Gaussian noise | variance $\in (0.0005 \sim 0.005)$ | 10 |
| Salt & paper noise | variance $\in (0.001 \sim 0.01)$ | 10 |
| Speckle noise | variance $\in (0.001 \sim 0.01)$ | 10 |
| | Blurring | |
| Gaussian blurring | filter size : $3, \sigma \in (0.5 \sim 5)$ | 10 |
| Circular blurring | radius $\in (0.2 \sim 2)$ | 10 |
| Motion blurring | len = $1, 2, 3\ \theta = 0°, 45°, 90°$ | 9 |
| | Geometric attacks | |
| Rotation | $\theta = \{\pm1°, \pm5°, \pm15°, \pm30°, \pm45°, \pm90°\}$ | 12 |
| Cropping & rotation | $\theta = \{\pm1°, \pm5°, \pm15°, \pm30°, \pm45°, \pm90°\}$ | 12 |
| Scaling | factor $\in (0.5 \sim 2.0)$ | 6 |
| JPEG compression | $QF \in (10 \sim 100)$ | 12 |

In order to show the identification performance of our DIBR 3D image hashing scheme is better than some other existing traditional 2D hashing schemes, two kinds of the current state-of-the-art 2D image hashing schemes are tested for experimental comparisons. One is the NMF-based hashing algorithm proposed in paper [21], and the other is the ring partition-based hashing algorithm proposed in [29, 30].

Suppose $\mathrm{IC} = \{\mathrm{IC}_i, 1 \le i \le N\}$ be the set of original center images. Then, we generate the compact hash $H(\mathrm{IC}_i)$ from each of the center images, and $H(\mathrm{IC}_i = h_1, h_2, \cdots, h_L)$ is the hash vector with length $L$ for center image $\mathrm{IC}_i$.

In this paper, we use correlation coefficient as the performance metric to evaluate the distance between two different hash vectors. Suppose $H(\mathrm{IC}_i)$ is the hash vector of one of the center image set, and $H(I_Q)$ is the query hash vector of distorted vision for either of the center image or their corresponding virtual images. Then, we calculate the correlation

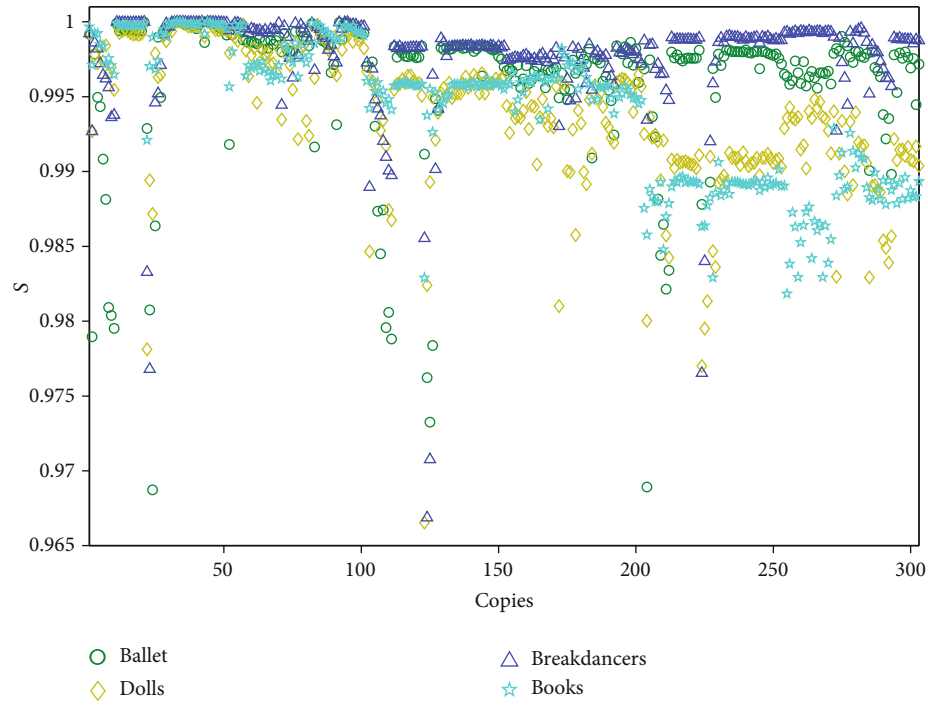


Figure 4: Robustness test based on four test images.

coefficient $S$ between $H(I_Q)$ and $H(\mathrm{IC}_i)$, and the query image is identified as the $i^{\mathrm{th}}$ original center image as

$$i = \arg\max_i \left\{ S\left(H\left(I_Q\right), H(\mathrm{IC}_i)\right)\right\}, \tag{9}$$

where $S(H(I_Q), H(\mathrm{IC}_i))$ is calculated as the correlation coefficient between $H(I_Q)$ and $H(\mathrm{IC}_i)$.

Higher identification accuracy rate means that the image attacked by common content-preserving operations can still be identified having similar perceptual content with the original one. When considering the problem of DIBR 3D image identification, high identification performance means that the virtual image should be identified having similar perceptual content with their corresponding center image even though the virtual images are attacked by common content-preserving operations.

As shown in Table 7, it is clear that the proposed hashing, NMF-based hashing, and ring partition-based hashing algorithms can achieve good identification performance, only taking into account the identification for center images. In [29, 30], they consider that all the perceptual distortions and malicious operations on digital images will not change the viewpoint, and the image center is usually unchanged, so it is relatively stable under geometric attacks such as rotation, scaling, and cropping after rotation. In fact, in the process of DIBR, the virtual image is generated from the center image through pixel shifting. Therefore, the hashing methods based on ring partition lose the advantage of generating robust hash for DIBR 3D image, as shown in Table 8. The experimental results show that the signal distorted virtual image can still be classified as the corresponding original center image with proposed hashing method. NMF-based method is sensitive to rotation attack due to the change of predefined position caused by geometric synchronization distortion. In contrast, the proposed hashing in this paper is robust to this kind of geometric attack. According to the experiment results listed in Table 9, it is clear that our DIBR 3D hashing scheme outperforms ring partition-based hashing schemes and NMF-based hashing scheme under content-preserving operations listed in above section.

Identification accuracy performances under combinational attacks between image geometric distortion attacks and signal distortion attacks are also tested with many images. As shown in Table 10, it is clear that the proposed hashing achieves good identification performances under most combinational attacks with slight degradations under Gaussian noise+geometric attacks (rotation with 45° and cropping after rotation with 45°) and speckle noise+geometric attacks (rotation with 45°). This means that our DIBR 3D hashing is robust against most of the above combinational attacks.

In this paper, FRR (false reject rate) and FAR (false accept rate) are also used to evaluate the perceptual robustness of proposed DIBR 3D image hashing scheme. FRR describes the error identification probability, the smaller FRR is, the better robustness of hash algorithm. FAR reflects the discrimination of hashing algorithm, the smaller FAR is, the better the discrimination. It is clear that an excellent hashing algorithm should have the minimum FRR and the minimum FAR with a certain threshold. As shown in Figure 7(d), for our hashing, the FRR and the FAR are zero with the

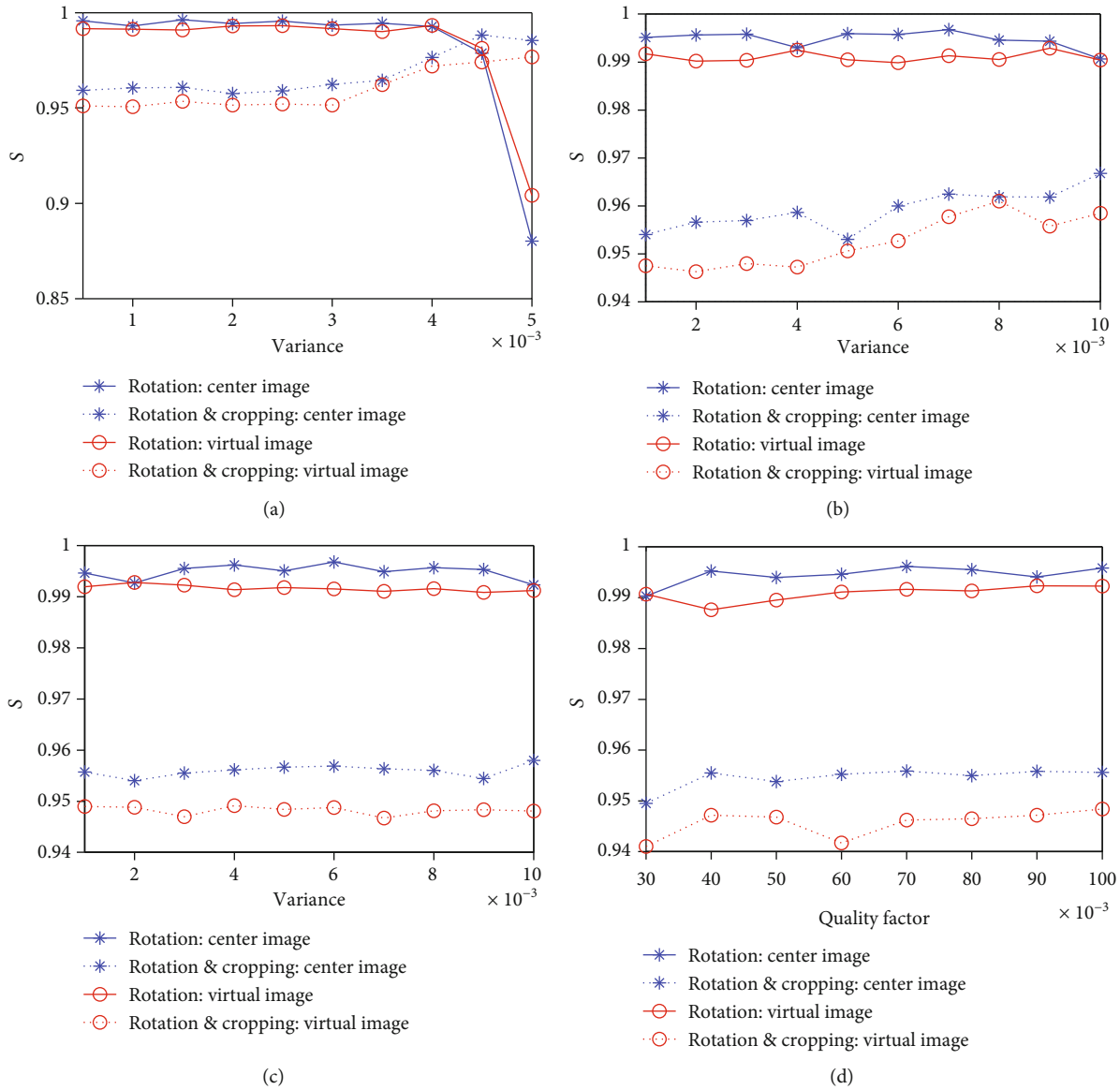


Figure 5: Our robustness performances under combinational attacks between rotation, cropping after rotation, and other operations. (a) Rotation+Gaussian noise and rotation and cropping+Gaussian noise. (b) Rotation+speckle noise and rotation and cropping+speckle noise. (c) Rotation+salt and paper noise and rotation and cropping+salt and paper noise. (d) Rotation+JPEG compression and rotation and cropping+JPEG compression.

threshold set from 0.86 to 0.93. This means that the proposed hashing could achieve the highest probability of true identification with zero false classification rate. As shown in Figure 7(a), for the NMF-based hashing [21], the minimum FRR and the minimum FAR are 0.164 when the threshold is set to 52. As shown in Figure 7(b), for the ring partition-based hashing [30], the minimum FRR and the minimum FAR are 0.176 when the threshold is set to 0.45. As shown in Figure 7(c), for the ring partition-based hashing [29], the minimum FRR and the minimum FAR are 0.16 when the threshold is set to 570. This experiment shows that the proposed hashing scheme is robust to common signal and geometric distortion attacks, such as additive noise, blurring, JPEG compression, scaling, and rotation.

The underlying reason is that these kinds of traditional 2D image hashing method consider that all perceptually insignificant distortions and malicious manipulations on a digital image would not lead to viewpoint changes, and the center of an image is generally preserved and thus relatively stable under geometric attacks such as rotation. In fact, virtual images are generated from center image with pixels shifting in the DIBR process. In paper [29, 30], they divide the image into several rings with the center of the image as the center. Using the pixels in every ring to form a secondary image, they extract the final hash from the secondary image. In the same way mentioned above, the different centers lead to form different secondary images, and the final hash vector of the center image is different from the hash vector of either virtual image.

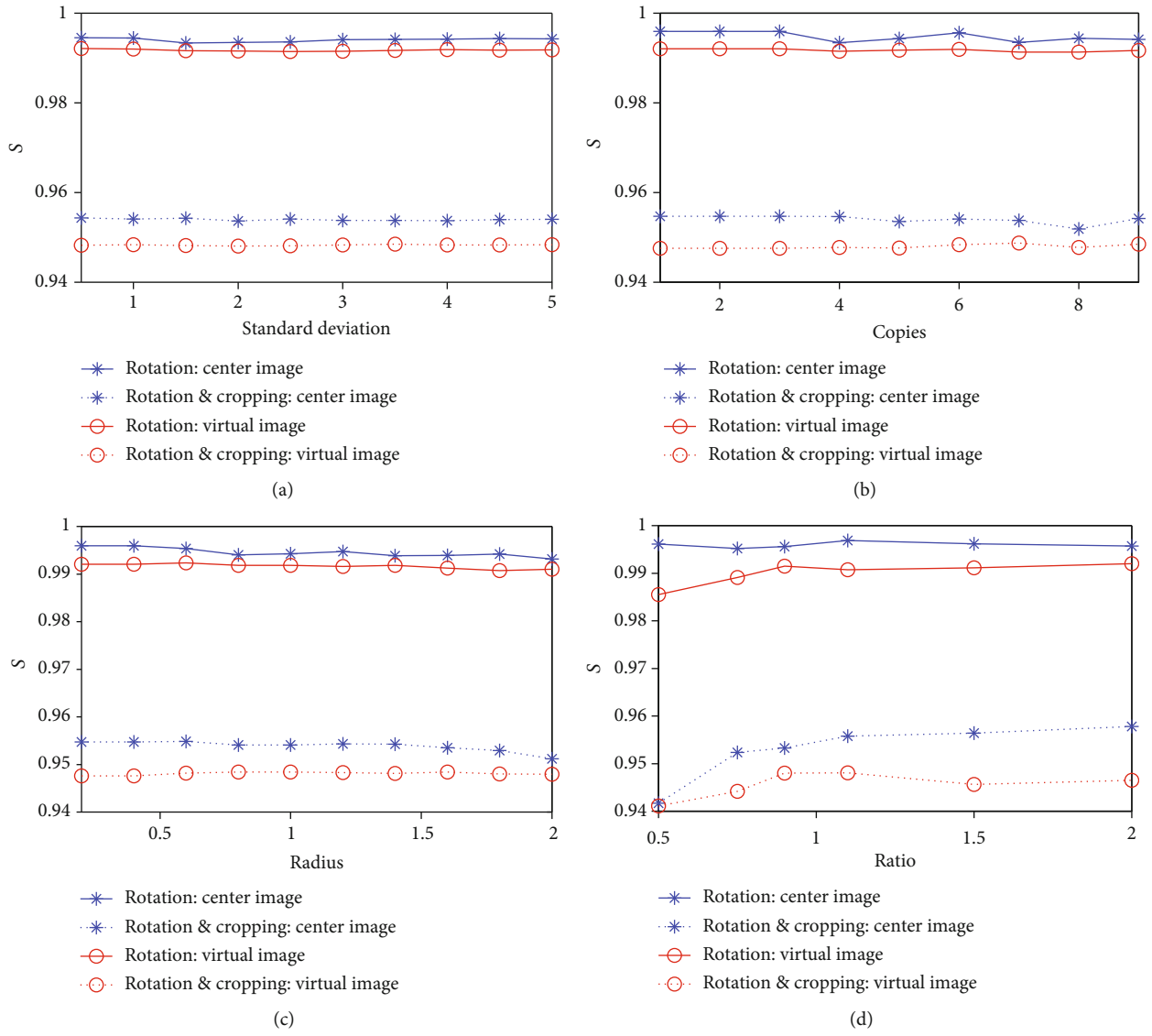


Figure 6: Our robustness performances under combinational attacks between rotation, cropping after rotation, and other operations. (a) Rotation+Gaussian blurring and rotation and cropping+Gaussian blurring. (b) Rotation motion blurring and rotation and cropping +motion blurring. (c) Rotation circular blurring and rotation and cropping+circular blurring. (d) Rotation scaling and rotation and cropping+scaling.

Table 6: The minimum perceptual distance under combinational attacks (rotation with different degrees).

| Combinational attacks | 2° | 10° | 45° |
|---|---|---|---|
| Noise+rotation | 0.9840 | 0.9284 | 0.8559 |
| Blurring+rotation | 0.9892 | 0.9907 | 0.9883 |
| JPEG compression+rotation | 0.9884 | 0.9887 | 0.9868 |
| Scaling+rotation | 0.9868 | 0.9831 | 0.9856 |
| Noise+cropping after rotation | 0.9827 | 0.9757 | 0.9470 |
| Blurring+cropping after rotation | 0.9891 | 0.9782 | 0.9476 |
| JPEG compression+cropping after rotation | 0.9845 | 0.9805 | 0.9410 |
| Scaling+cropping after rotation | 0.9852 | 0.9751 | 0.9412 |

Table 7: Identification accuracy performances for center image by different methods.

| Manipulation | Our | [30] | [29] | [21] |
|---|---|---|---|---|
| Additive noise | | | | |
| Gaussian noise | 100% | 97.78% | 100% | 100% |
| Salt & paper noise | 100% | 100% | 100% | 100% |
| Speckle noise | 100% | 100% | 100% | 100% |
| Blurring | | | | |
| Gaussian blurring | 100% | 100% | 100% | 100% |
| Circular blurring | 100% | 100% | 100% | 100% |
| Motion blurring | 100% | 100% | 100% | 100% |
| Geometric attacks | | | | |
| Rotation | 100% | 100% | 100% | 34.26% |
| Cropping & rotation | 100% | 100% | 100% | 38.89% |
| Scaling | 100% | 100% | 100% | 100% |
| JPEG compression | 100% | 99.07% | 100% | 100% |

Table 8: Identification accuracy performances for virtual image by different methods.

| Manipulation | Our | [30] | [29] | [21] |
|---|---|---|---|---|
| Additive noise | | | | |
| Gaussian noise | 100% | 69.45% | 65.56% | 80.00% |
| Salt & paper noise | 100% | 72.22% | 68.34% | 82.78% |
| Speckle noise | 100% | 70.00% | 67.78% | 84.44% |
| Blurring | | | | |
| Gaussian blurring | 100% | 71.67% | 65.00% | 87.77% |
| Circular blurring | 100% | 71.11% | 63.89% | 90.55% |
| Motion blurring | 100% | 69.76% | 67.90% | 88.89% |
| Geometric attacks | | | | |
| Rotation | 100% | 55.56% | 50.00% | 32.87% |
| Cropping & rotation | 100% | 56.02% | 49.54% | 37.04% |
| Scaling | 100% | 68.98% | 71.30% | 87.96% |
| JPEG compression | 100% | 70.84% | 68.34% | 84.72% |

*5.3. Robustness against Baseline Distance Adjustment.* As shown in Section 3, in the DIBR process, a virtual image can be generated using an appropriate baseline distance of $t_x$. Usually, $t_x$ is set different to suit different people's vision. Because $t_x$ is not fixed during DIBR rendering, baseline distance adjustment may affect the identification performance of virtual image. In order to show the robustness of the proposed hash method for adjusting the baseline distance, the range of the baseline distance $t_x$ is from 5% to 7% of the image width. As shown in Table 11, the identification accuracy of different baseline distance is invariable.

*5.4. Key Dependence.* To enhance the security of hashing scheme, a secret key is usually used in the processes of feature extraction and feature compression to generate the final hash. As a result, the key-based hashing scheme is key dependent, making the hash unpredictable to prevent unauthorized access.

Table 9: Identification accuracy performances for center and virtual image by different methods.

| Manipulation | Our | [30] | [29] | [21] |
|---|---|---|---|---|
| Additive noise | | | | |
| Gaussian noise | 100% | 78.89% | 77.04% | 86.67% |
| Salt & paper noise | 100% | 81.48% | 78.89% | 88.52% |
| Speckle noise | 100% | 80.00% | 78.52% | 89.63% |
| Blurring | | | | |
| Gaussian blurring | 100% | 81.11% | 76.67% | 91.85% |
| Circular blurring | 100% | 80.74% | 75.93% | 93.70% |
| Motion blurring | 100% | 79.84% | 78.60% | 92.59% |
| Geometric attacks | | | | |
| Rotation | 100% | 70.37% | 66.67% | 33.33% |
| Cropping & rotation | 100% | 70.68% | 66.36% | 37.65% |
| Scaling | 100% | 79.32% | 80.87% | 91.98% |
| JPEG compression | 100% | 80.25% | 78.89% | 89.91% |

Table 10: Our identification accuracy performances for center image and virtual image under combinational attacks (rotation with different degrees).

| Combinational attacks | 2° | 10° | 45° |
|---|---|---|---|
| Gaussian noise+rotation | 100% | 100% | 95.19% |
| Salt & paper noise+rotation | 100% | 100% | 100% |
| Speckle noise+rotation | 100% | 100% | 95.93% |
| Gaussian blurring+rotation | 100% | 100% | 100% |
| Circular blurring+rotation | 100% | 100% | 100% |
| Motion blurring+rotation | 100% | 100% | 100% |
| JPEG compression+rotation | 100% | 100% | 100% |
| Scaling+rotation | 100% | 100% | 100% |
| Gaussian noise+cropping after rotation | 100% | 100% | 98.89% |
| Salt & paper noise+cropping after rotation | 100% | 100% | 100% |
| Speckle noise+cropping after rotation | 100% | 100% | 100% |
| Gaussian blurring+cropping after rotation | 100% | 100% | 100% |
| Circular blurring+cropping after rotation | 100% | 100% | 100% |
| Motion blurring+cropping after rotation | 100% | 100% | 100% |
| JPEG compression+cropping after rotation | 100% | 100% | 100% |
| Scaling+cropping after rotation | 100% | 100% | 100% |

In the proposed hashing scheme, only pixels with $M$ different gray levels are used to construct the secondary image. Using a key-based sequence $P(M)$ to select pixel groups, the security of proposed hashing scheme is enhanced. To validate key dependence of proposed hashing scheme, four images "Breakdancers," "Ballet," "Dolls," and "Books" are adopted.

For each image, hashes are generated with 100 different keys. Then, we calculate the correlation coefficients between the original key-based hash and hashes with different keys; it can be found that all correlation coefficients between different hashes of the four images are smaller. It should be noted that the parameters of hash generation are kept unchanged

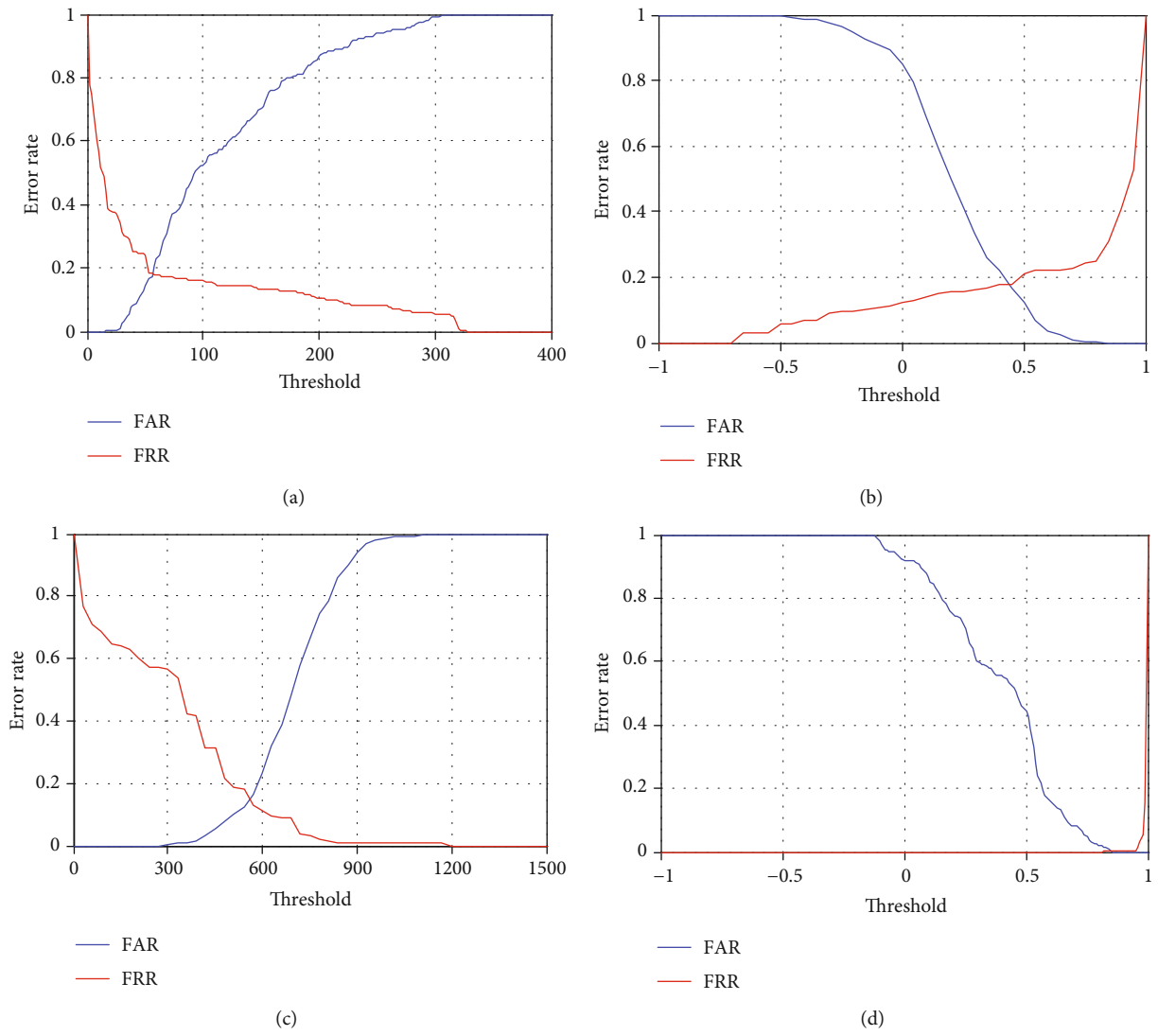


Figure 7: (a) The FAR and FRR of hash algorithm in [21]. (b) The FAR and FRR of hash algorithm in [30]. (c) The FAR and FRR of hash algorithm in [29]. (d) The FAR and FRR of proposed hash algorithm.

Table 11: Identification accuracy performances by proposed method with different baseline distances.

| Manipulation | 5% | 6% | 7% |
|---|---|---|---|
| Additive noise | | | |
| Gaussian noise | 100% | 100% | 100% |
| Salt & paper noise | 100% | 100% | 100% |
| Speckle noise | 100% | 100% | 100% |
| Blurring | | | |
| Gaussian blurring | 100% | 100% | 100% |
| Circular blurring | 100% | 100% | 100% |
| Motion blurring | 100% | 100% | 100% |
| Geometric attacks | | | |
| Rotation | 100% | 100% | 100% |
| Cropping & rotation | 100% | 100% | 100% |
| Scaling | 100% | 100% | 100% |
| JPEG compression | 100% | 100% | 100% |

except the key-based sequence $P(M)$ for selecting pixel groups in this experiment. Then, the correlation coefficients between the original key-based hash and other 100 hashes with different keys are computed for the four images mentioned above, and the obtained results are illustrated in Figure 8, where the $x$-axis is the index of key and the $y$-axis is the correlation coefficient $S$, which represents the hash distance. For the image of "Breakdancers," the maximum, the minimum, and the average distances are 0.4507, -0.1849, and 0.1525, respectively. For the image of "Ballet," the maximum, the minimum, and the average distances are 0.4754, 0.0838, and 0.3185, respectively. For the image of "Dolls," the maximum, the minimum, and the average distances are 0.3162, -0.2067, and 0.1226, respectively. For the image of "Books," the maximum, the minimum, and the average distances are 0.5470, -0.0440, and 0.3319, respectively. It is clear that the maximum distances between the original key-based hash and other 400 hashes with different keys are lower than 0.96. This experimental result shows that the security of

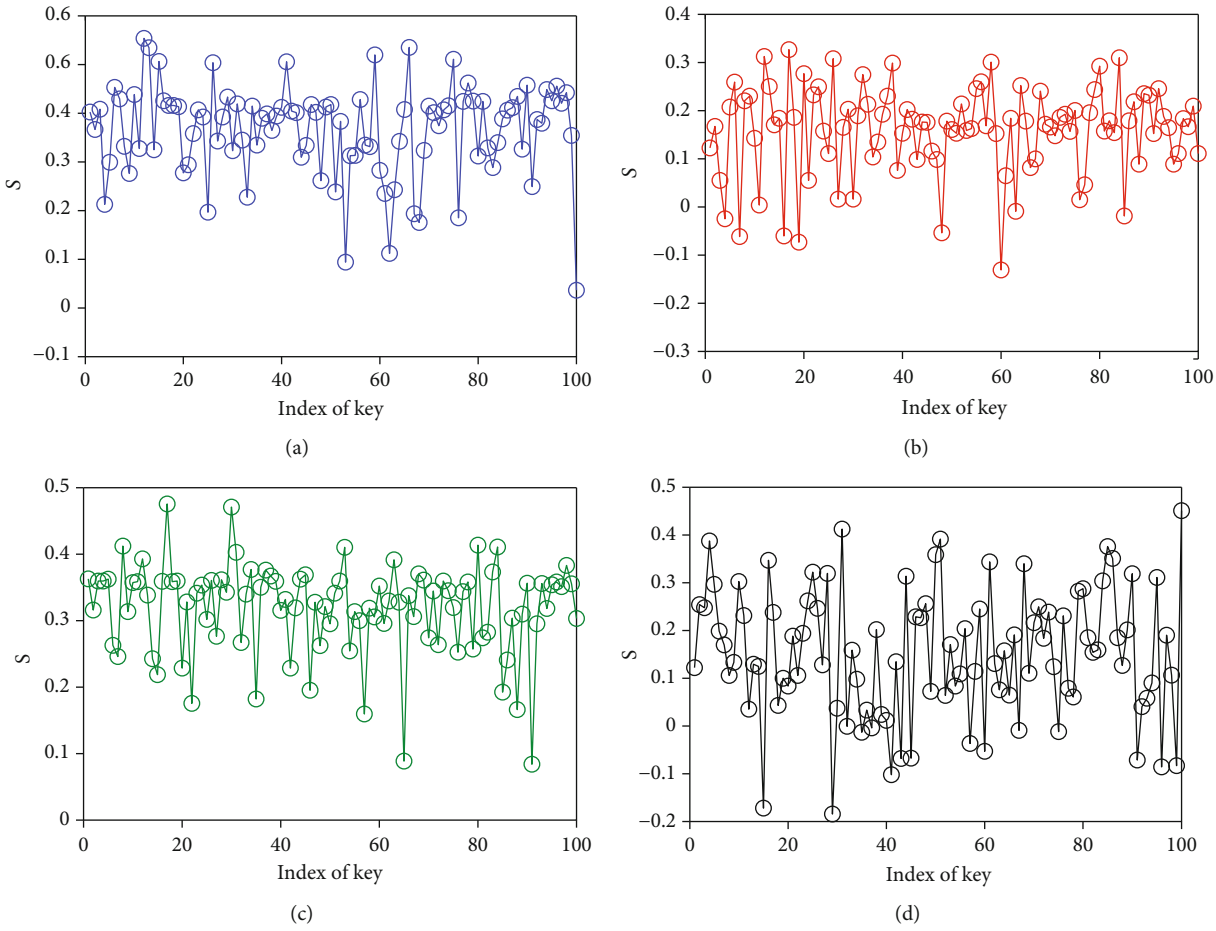


Figure 8: (a) Correlation coefficients between hashes of "Books" generated by different keys, (b) correlation coefficients between hashes of "Dolls" generated by different keys, (c) correlation coefficients between hashes of "ballet" generated by different keys, and (d) correlation coefficients between hashes of "Breakdancers" generated by different keys.

proposed hashing scheme is enhanced with a key-based sequence $P(M)$ to select pixel groups.

## 6. Conclusions

In this paper, we propose a pixel grouping and NMF-based DIBR 3D image hashing scheme, which can be used for virtual image identification and retrieval. Low-pass filtering and histogram shape-based pixel grouping are the key steps to make proposed hashing scheme robust to common content-preserving manipulations, and the approximate invariance of histogram shape to cropping and DIBR operations ensures that our DIBR 3D image hashing scheme also has better performance for virtual image identification. The experiment results have shown that the proposed DIBR 3D image hashing resists to common content-preserving manipulations including signal distortion attacks and geometric distortion attacks. However, the proposed hashing method may identify an input image with different content to be visually identical, when the input image has the same histogram shape. We will solve this problem in the future work.

## Data Availability

To get the dataset for discrimination, please visit http://www.cs.washington.edu/research/imagedatabase/groundtruth/. Further details can be provided upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

We would like to thank the anonymous reviewers for their helpful comments and suggestions, and their comments and suggestions help us to improve the quality of this paper. This work is supported by the National Natural Science Foundation of China (Grant Number: 61702224), the Special Funds of Heilongjiang University of the Fundamental Research Funds for the Heilongjiang Province (RCCXYJ201811 and RCCXYJ201812), the Open Fund of the State Key Laboratory of Information Security (2019-ZD-05), the Natural Science Foundation of Zhejiang Province (No. LY18F020020),

the Guangxi Key Laboratory of Cryptography and Information Security (No. GCIS201904), and the Heilongjiang Provincial Natural Science Foundation of China (Grant No. LH2020F044).

## References

[1] C. Fehn, "Depth-image-based rendering (DIBR) compression and transmission for a new approach on 3D-TV," in *Proceedings of the SPIE Stereoscopic Displays and Virtual Reality Systems XI*, pp. 93–104, San Jose, CA, USA, May 2004.

[2] A. Gionis, P. Indyky, and R. Motwani, "Similarity search in high dimensions via hashing," in *The 25th VLDB Conference*, pp. 518–529, Edinburgh, Scotland, 1999.

[3] K. Li, G. Qi, J. Ye, and K. A. Hua, "Linear subspace ranking hashing for cross-modal retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 9, pp. 1825–1838, 2017.

[4] X. Wang, T. Zhang, G. Qi, J. Tang, and J. Wang, "Supervised quantization for similarity search," in *2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2018–2026, Las Vegas, NV, USA, June 2016.

[5] C. Lu, S. C. Y. Hsu, S. W. Sun, and P. C. Chang, "Robust mesh based hashing for copy detection and tracing of images," in *2004 IEEE Conference on Multimedia and Expo*, pp. 731–734, Taipei, Taiwan, June 2004.

[6] X. Zhu, X. Li, S. Zhang, Z. Xu, L. Yu, and C. Wang, "Graph PCA hashing for similarity search," *IEEE Transactions on Multimedia*, vol. 19, no. 9, pp. 2033–2044, 2017.

[7] X. Lv and Z. J. Wang, "Perceptual image hashing based on shape contexts and local feature points," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 1081–1093, 2012.

[8] C. Qin, M. Sun, and C. C. Chang, "Perceptual hashing for color images based on hybrid extraction of structural features," *Signal Processing*, vol. 142, pp. 194–205, 2017.

[9] J. Song, Y. Yang, X. Li, Z. Huang, and Y. Yang, "Robust hashing with local models for approximate similarity search," *IEEE Transactions on Cybernetics*, vol. 44, no. 7, pp. 1225–1236, 2014.

[10] X. Lu, X. Zheng, and X. Li, "Latent semantic minimal hashing for image retrieval," *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 355–368, 2017.

[11] F. Ahmed, M. Y. Siyal, and V. U. Abbas, "A secure and robust hash-based scheme for image authentication," *Signal Processing*, vol. 90, no. 5, pp. 1456–1470, 2010.

[12] C. Wang, D. Wang, Y. Tu, G. Xu, and H. X. Wang, "Understanding node capture attacks in user authentication schemes for wireless sensor networks," *IEEE Transactions on Dependable and Secure Computing*, 2020.

[13] D. Wang, W. T. Li, and P. Wang, "Measuring two-factor authentication schemes for real-time data access in industrial wireless sensor networks," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 9, pp. 4081–4092, 2018.

[14] S. M. Qiu, D. Wang, G. A. Xu, and S. Kumari, "Practical and provably secure three-factor authentication protocol based on extended chaotic-maps for mobile lightweight devices," *IEEE Transactions on Dependable and Secure Computing*, 2020.

[15] X. Lv and Z. J. Wang, "Reduced-reference image quality assessment based on perceptual image hashing," in *2009 IEEE Conference on Image Processing*, pp. 4361–4364, Cairo, Egypt, November 2009.

[16] W. Lu and M. Wu, "Multimedia forensic hash based on visual words," in *2010 IEEE Conference on Image Processing*, pp. 989–992, Hong Kong, China, September 2010.

[17] C. Yan, C. Pun, and X. Yuan, "Multi-scale image hashing using adaptive local feature extraction for robust tampering detection," *Signal Processing*, vol. 121, pp. 1–16, 2016.

[18] Z. Tang, Y. Dai, and X. Zhang, "Perceptual hashing for color images using invariant moments," *Applied Mathematics & Information Sciences*, vol. 6, no. 2S, pp. 643–650, 2012.

[19] V. Monga and B. L. Evans, "Perceptual image hashing via feature points: performance evaluation and tradeoffs," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3452–3465, 2006.

[20] S. Kozat, R. Venkatesan, and M. Mihcak, "Robust perceptual image hashing via matrix invariants," in *2004 International Conference on Image Processing*, pp. 3443–3446, Singapore, Singapore, October 2004.

[21] V. Monga, "Robust and secure image hashing via non-negative matrix factorizations," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 376–390, 2007.

[22] Z. Tang, Z. Huang, X. Zhang, and H. Lao, "Robust image hashing with multidimensional scaling," *Signal Processing*, vol. 137, pp. 240–250, 2017.

[23] S. Roy and Q. Sun, "Robust hash for detecting and localizing image tampering," in *2007 IEEE International Conference on Image Processing*, pp. 117–120, San Antonio, TX, USA, September-October 2007.

[24] Y. Lei, Y. Wang, and J. Huang, "Robust image hash in Radon transform domain for authentication," *Signal Processing: Image Communication*, vol. 26, no. 6, pp. 280–288, 2011.

[25] Y. Li, Z. Lu, C. Zhu, and X. Niu, "Robust image hashing based on random Gabor filtering and dithered lattice vector quantization," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1963–1980, 2012.

[26] Z. Tang, S. Wang, X. Zhang, and W. Wei, "Structural feature-based image hashing and similarity metric for tampering detection," *Fundamenta Informaticae*, vol. 106, no. 1, pp. 75–91, 2011.

[27] C. Qin, X. Chen, X. Luo, X. Zhang, and X. Sun, "Perceptual image hashing via dual-cross pattern encoding and salient structure detection," *Information Sciences*, vol. 423, pp. 284–302, 2018.

[28] Z. Tang, L. Chen, X. Zhang, and S. Zhang, "Robust image hashing with tensor decomposition," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 3, pp. 549–560, 2019.

[29] Z. Tang, X. Zhang, X. Li, and S. Chao, "Robust image hashing with ring partition and invariant vector distance," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 1, pp. 200–214, 2016.

[30] Z. Tang, X. Zhang, and S. Chao, "Robust perceptual image hashing based on ring partition and NMF," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 3, pp. 711–724, 2014.

[31] L. Zhang and W. J. Tam, "Stereoscopic image generation based on depth images for 3d TV," *IEEE Transactions on Broadcasting*, vol. 51, no. 2, pp. 191–199, 2005.

[32] S. Xiang, H. J. Kim, and J. Huang, "Invariant image watermarking based on statistical features in the low-frequency

domain," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 6, pp. 777–790, 2008.

[33] T. Zong, Y. Xiang, I. Natgunanathan, S. Guo, W. Zhou, and G. Beliakov, "Robust histogram shape-based method for image watermarking," *IEEE Transactions on Circuits & Systems for Video Technology*, vol. 25, no. 5, pp. 717–729, 2015.

[34] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, Minneapolis, MN, USA, June 2007.

[35] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2007.

[36] Ground Truth DatabaseMay 2008, http://www.cs.washington.edu/research/imagedatabase/groundtruth/.