Pixel Grouping Based Image Hashing for DIBR 3D Image

Chen Cui[1], Xujun Wu[1], Jun Yang[2], and Juyan Li[1](✉)

[1] School of Information Science and Technology, Heilongjiang University, Harbin, Heilongjiang, China
2018012@hlju.edu.cn, 1149347287@qq.com, lijuyan587@163.com

[2] College of Mathematics Physics and Information Engineering, Jiaxing University, Jiaxing 314000, China
yangj95@mail2.sysu.edu.cn

Abstract. Most of the traditional 2D image hashing schemes do not take into account the change of viewpoint to construct the hash vector, resulting in the classification accuracy rate is unsatisfactory when applied in identification for Depth-image-based rendering (DBIR) 3D image. In this work, pixel grouping according to histogram shape and Nonnegative matrix factorization (NMF) is applied to design DIBR 3D image hashing with better robustness resist to geometric distortions and higher classification accuracy rate for virtual images identification. Experiments show that the proposed hashing is robust to common signal and geometric distortion attacks, such as additive noise, blurring, JPEG compression, scaling and rotation. When compared with the state-of-art schemes for traditional 2D image hashing, the proposed hashing provides better performances under above distortion attacks when considering the virtual images identification.

Keywords: DIBR 3D image identification · Image hashing · Histogram · Nonnegative matrix factorization (NMF)

1 Introduction

Depth-image-based Rendering (DIBR) [1] is a kind of 3D representation technology, by which virtual right image and right image are generated from the center image according to the depth information represented 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

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, RCCXYJ201812). The Open Fund of the State Key Laboratory of Information Security (2019-ZD-05). The Natural Science Foundation of Zhejiang Province (No. LY18F020020).

D. Wang et al. (Eds.): SPNCE 2020, LNICST 344, pp. 356–369, 2021.
https://doi.org/10.1007/978-3-030-66922-5_25

depth-image-based Rendering operation to generated virtual image pair for 3D video. As a matter of fact, either of the center image, the virtual left image and virtual right image may suffer from illegal or unauthorized re-distributions. In order to resolve this problem, robust perceptual hashing has been widely used for digital multimedia protection. As variety of copies for the center image and virtual images existing, image hashing can also help us to find the similar one and detect the tempered. In this paper, we focus on designing a robust image hashing scheme for DIBR 3D image identification.

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

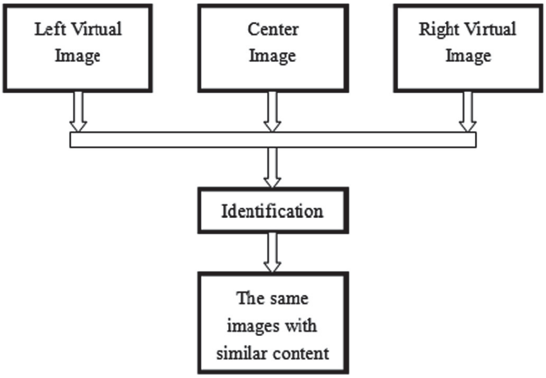

Fig. 1. The character of image hashing for DIBR 3D images

2 Related Work

Robust images hashing has been extensively studied for content-based identification of traditional 2D images. As feature extraction affects the identification performance for image hashing, many existing methods focus on extracting robust features resist to content-preserving operations [2,3]. In addition, some matrix analysis methods have been adopted to extract the robust features for hash generation, such as singular value decomposition (SVD) [4] and non-negative matrix factorization (NMF) [5], which are robust to most kinds of signal distortion attacks but sensitive to geometric distortions such as rotation.

In order to make the hashing scheme robust to geometric distortion attacks such as rotation, some robust image hashing schemes are proposed to extract the geometric-invariant features for generating the final hash vector [6]. Lv et al. [7] proposed a shape contexts and local feature points based image hashing scheme. Compressing the descriptors of stable 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. Tang et al. [8,9] proposed a kind of robust image hashing scheme based on ring partition. Using the pixels in every ring to form a secondary image resist 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 dose not change, 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. In fact, the center of center image and virtual images are different, which is caused by DIBR 3D process. As a result, this kind of traditional 2D hashing scheme would not show good performance when applied for DIBR 3D image identification.

Some of the state-of-art traditional 2D image hashing schemes which resist to geometric distortions do not take into account the situation about viewpoint changing [7,8]. 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 images identification. The key contribution is using the approximate invariance of histogram shape to eliminate as much difference between center image and virtual image as possible. The rest of this paper is organized as follows: Sect. 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.

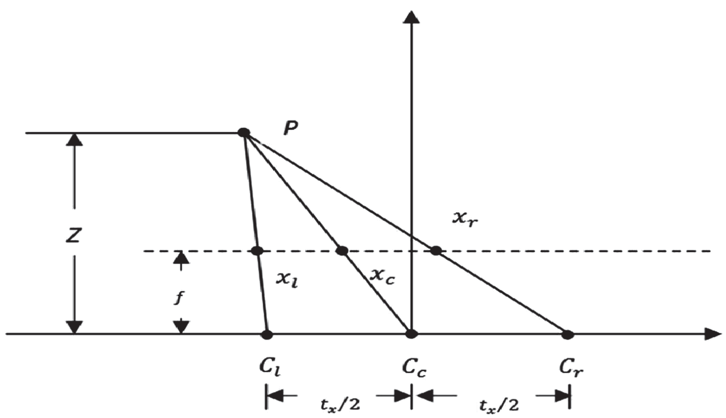

Fig. 2. The relationship of pixel in the left image, center image and right image

3 Review of Depth-Image-Based Rendering Process

Figure 2 illustrates the relationship between the center image and virtual images generated by DIBR operations [10]. Suppose P is a point in the space, C_c, C_l and C_r represent the center viewpoint, left viewpoint and the right viewpoint, respectively, f represents the focal length of the center viewpoint, 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 Fig. 2, x-coordinate of pixel in virtual images is computed as

$$\begin{aligned} x_l &= x_c + \tfrac{t_x}{2}\tfrac{f}{Z}, \\ x_r &= x_c - \tfrac{t_x}{2}\tfrac{f}{Z}, \end{aligned} \tag{1}$$

$$Z(v) = Z_{far} + v \times \tfrac{Z_{near} - Z_{far}}{255}, v \in [0, 255] \tag{2}$$

Where x_l, x_c and x_r represent the x–coordinate of pixels in the left virtual image, center image and virtual right image, respectively. In fact, the gray values of pixels in gray image are not the real depth value. Pixel with gray value close to 255 indicates that P is close to the near clipping plane Z_{near}. On the other hand, pixel with gray value close to 0 indicates that P is close to the far clipping plane Z_{far}. According to formula 2, the depth value $Z(v)$ of P is computed, where v represents the gray value.

4 Proposed Image Hashing

Our DIBR 3D image hashing algorithm 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 an 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 Pre-processing

Low-pass filtering is adopted to extract the low-frequency component of original center image, aiming to enhance robustness of proposed hashing scheme to some common content-preserving manipulations [13]. The low-frequency component IC_{low} of original center image IC is obtained as

$$IC_{low}(x, y) = G(x, y, \sigma) * I(x, y) \tag{3}$$

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

$$G(x,y,\sigma)=\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{4}$$

σ is the standard difference. According to parameters setting in [13], σ 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, aiming to ensure the security of proposed hashing algorithm. With a key-based sequence $P(M)=\{p_i|i=1...M, 0\leq p_i\leq 255\}$, M gray levels $h_1,h_2,...,h_M$ are selected for pixels grouping, where $h_i=p_i$. The set of selected gray level is represented as

$$H_M=\{h_i|i=1,....,M\} \tag{5}$$

After resizing I_{low} to $m\times m$, pixels with L_B neighboring gray levels in H_M are selected to form one pixel group. In total, $n=\lfloor M/L_B\rfloor$ groups are formed, where $\lfloor\bullet\rfloor$ 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 modeled as

$$V=[v_1,v_2,v_3,...,v_n] \tag{6}$$

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

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 a 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 similarity between two image hash vectors $Hash1$ and $Hash2$. The correlation coefficient $S(Hash1,Hash2)$ is defined as

$$S(Hash1,Hash2)=\frac{cov(Hash1,Hash2)}{\sqrt{D(Hahs1)}\sqrt{D(Hahs1)}} \tag{7}$$

According to formula 7, S ranges from -1 to 1, a bigger S value indicates that the input image is more similar with the original corresponding center image. If the correlation coefficient S is higher than the threshold predefined, the input image is viewed as perceptual content unchanged. If the correlation coefficient S 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 value when computing the perceptual distance from their corresponding original center image. According to experiment results listed in Table 1 and Table 2, some virtual images are viewed different from the original center image when the hashing method proposed in [9] is adopted. It is clearly that our DIBR 3D image hashing scheme can identify the virtual images with visual contents the same as the original center one.

Table 1. Perceptual distance between center image and left virtual image computed by different hashing method.

Image	Proposed method	Method in [9]
Breakdancers	0.9984	−0.3817
Dolls	0.9952	−0.7150
Books	0.9982	−0.7499
ballet	0.9984	0.8183

Table 2. Perceptual distance between center image and right virtual image computed by different hashing method.

Image	Proposed method	Method in [9]
Breakdancers	0.9990	0.5647
Dolls	0.9909	0.7812
Books	0.9982	0.8849
ballet	0.9980	0.9093

5 Experimental Results

120 different color images are collected from the Ground Truth Database [14] in order to test discriminative capability of proposed hashing. The hash vectors are generated for these 120 images, then 7140 correlation coefficients 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% pairs of different images are identified with the similar content. 0.09% pairs of different images are identified with the similar content with T is set as 0.94. No pair of different images is identified with the similar content when T is set as 0.98.

Table 3. Content-preserving operations and the parameters setting.

Manipulation	Parameters 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	$\theta \in (-10° \sim +10°)$	12
Cropping & Rotation	$\theta \in (-10° \sim +10°)$	12
Scaling	$factor \in (0.5 \sim 2.0)$	6
JPEG compression	$QF \in (10 \sim 100)$	10

Database with 2727 images is constructed to evaluate the identification performance for DIBR 3D image. In this database, we select 9 pairs of center and depth images from Middlebury Stereo Datasets [11] and Microsoft Research 3D Video Datasets [12], and sizes of these images are ranging from 450×375 to 1390×1110. Hashes extracted from center image, virtual left image, 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 implemented these 10 classes 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 3.

Firstly, four pairs of the center and depth image are selected from above dataset. They are Breakdancers, Books, Dolls, and ballet as listed in Table 1. each virtual images pair and the center image are attacked by the content-preserving operations listed in Table 3. As shown in Fig. 3, no pair of visually identical images (including the distorted center and virtual images) is identified as different content when the threshold T is set to 0.96.

In order to show the identification performance of our DIBR 3D image hashing scheme is better than some other existing traditional 2D hashing schemes,

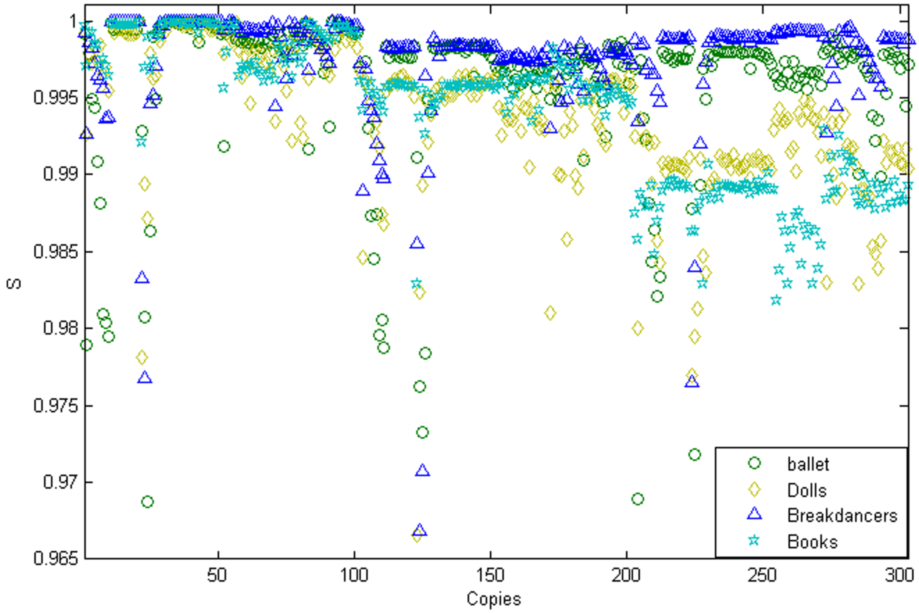

Fig. 3. Robustness test based on four test images.

two kinds of the current state-of-the-art 2D image hashing schemes are selected to for experimental comparisons. One is the Shape Contexts and Local Feature Points based hashing algorithm (RSCH) proposed in paper [7], and the other is the Ring Partition and NMF based hashing algorithm proposed in [9].

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

In this paper, we use correlation coefficient as the performance metric to evaluate the distance between two different hash vectors $H(I_1)$ and $H(I_2)$. Suppose $H(IC_i)$ 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 coefficient S between $H(I_Q)$ and $H(IC_i)$, and the query image is identified as the i-th original center image as

$$i = \underset{i}{argmax} \{S(H(I_Q), H(IC_i))\} \tag{8}$$

Where $S(H(I_Q), H(IC_i))$ is calculated as the correlation coefficient between $H(I_Q)$ and $H(IC_i)$.

Higher identification accuracy rate means that the images attacked by common content-preserving operations can still be identified having similar perceptual content with the original one, no matter attacked by common content-preserving operations. When considering the DIBR 3D image identification, high identification performance means that the virtual images should be identified

Table 4. Identification accuracy performances for center and virtual images by different methods.

Manipulation	Our method	Method in [9]	Method in [7]
Additive Noise			
Gaussian Noise	100%	78.89%	69.26%
Salt & Paper Noise	100%	81.48%	78.52%
Speckle Noise	100%	80.00%	77.78%
Blurring			
Gaussian Blurring	100%	81.11%	71.48%
Circular Blurring	100%	80.74%	78.14%
Motion Blurring	100%	79.84%	77.41%
Geometric Attacks			
Rotation	100%	73.46%	61.72%
Cropping & Rotation	100%	74.07%	71.91%
Scaling	100%	79.32%	64.20%
JPEG Compression	100%	80.25%	84.88%

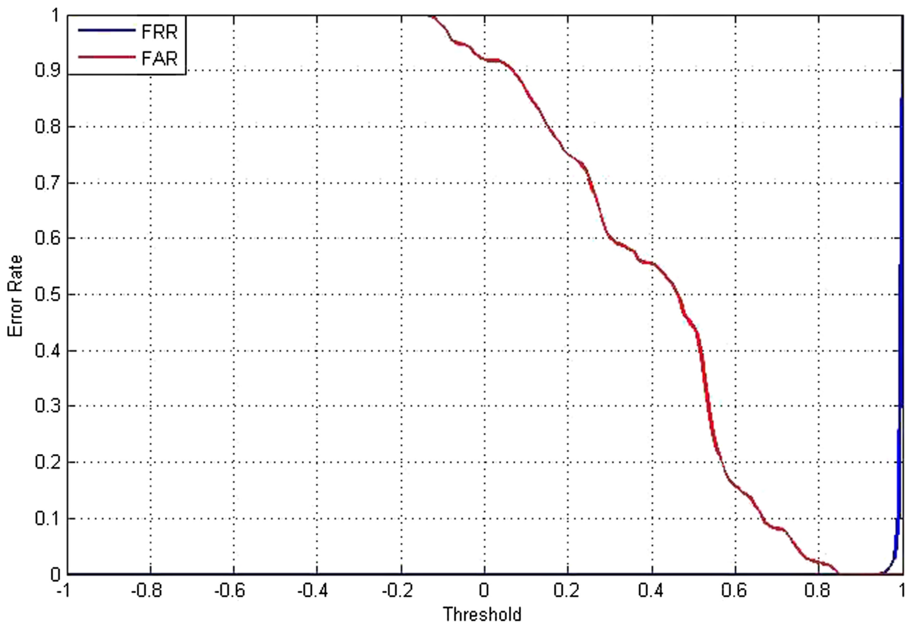

Fig. 4. Robustness test based on four test images.

having similar perceptual content with their corresponding center image even though the virtual images are attacked by common content-preserving operations.

In this paper, FRR(false reject rate) and FAR(false accept rate) are used to evaluate the 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 clearly that an excellent hashing algorithm should have the minimum FRR and minimum FAR with a certain threshold. As shown in Fig. 4, the FRR and FAR are zero when the threshold is set from 0.86 to 0.93. 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.

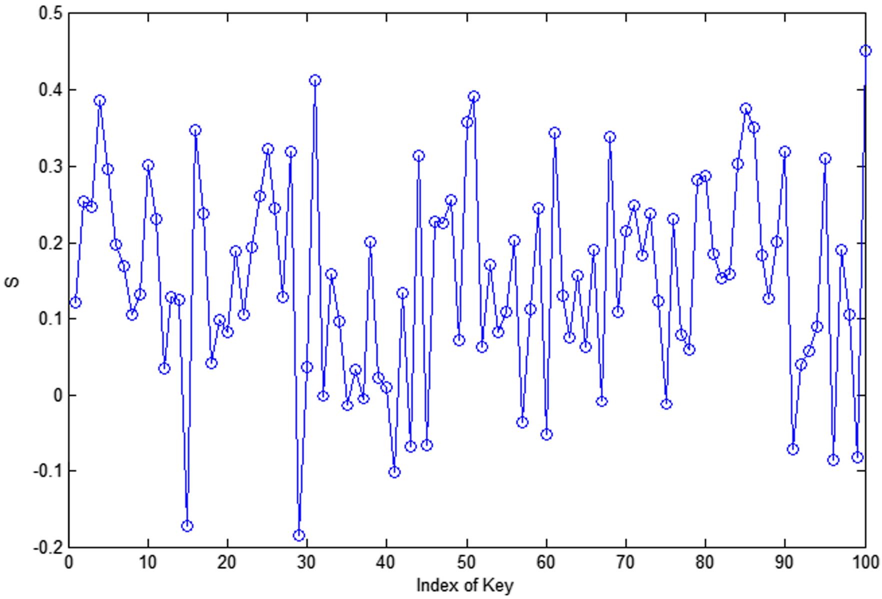

Fig. 5. Correlation coefficients between hashes of "Breakdancers" generated by different keys.

According to the experiment results listed in Table 4, it is clearly that our DIBR 3D hashing scheme outperforms Ring partition hashing scheme and RSCH hashing scheme under content-preserving operations listed in above section. 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

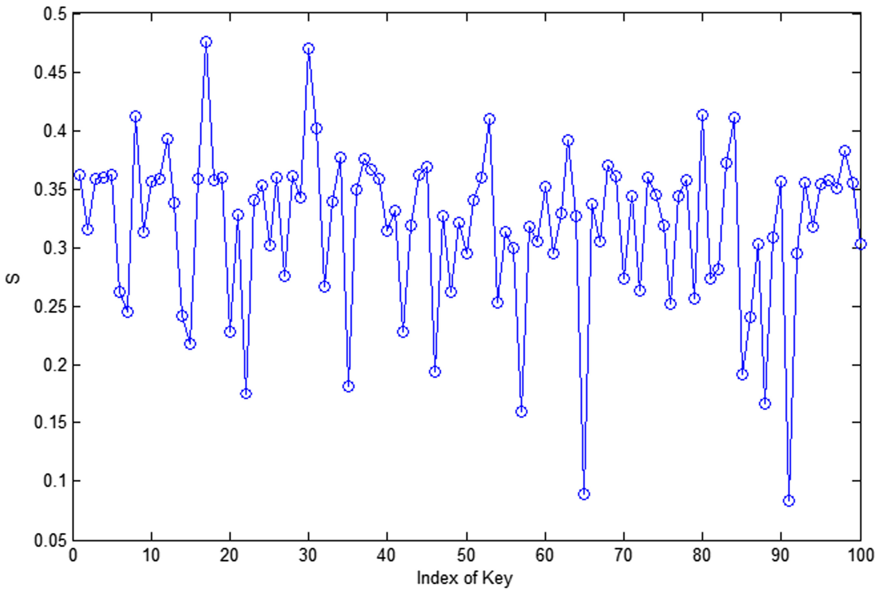

Fig. 6. Correlation coefficients between hashes of "Ballet" generated by different keys.

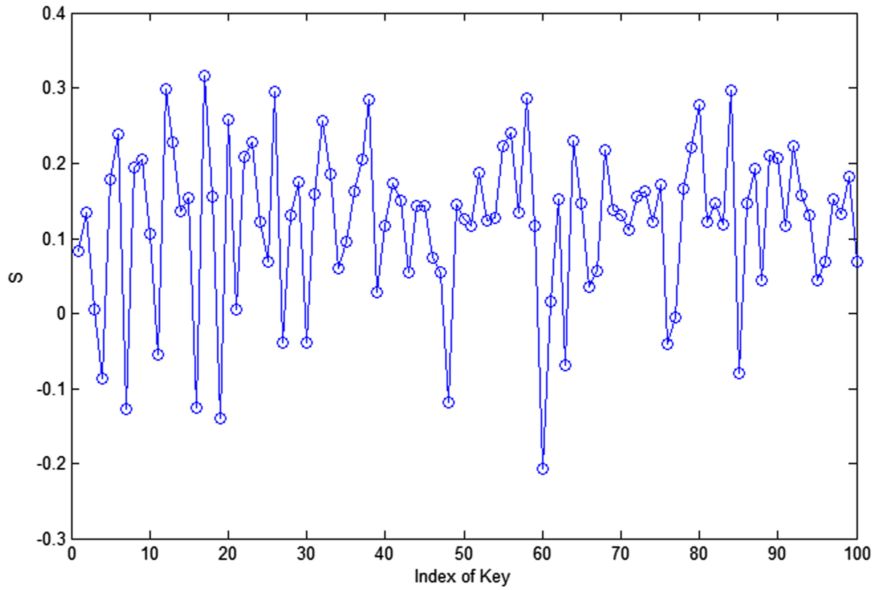

Fig. 7. Correlation coefficients between hashes of "Dolls" generated by different keys.

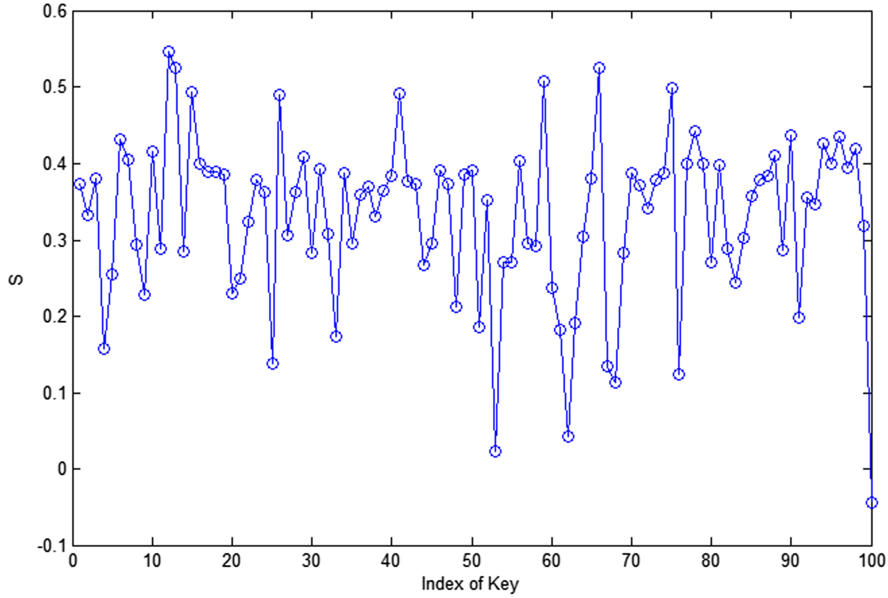

Fig. 8. Correlation coefficients between hashes of "Books" generated by different keys.

attacks such as rotation. In fact, virtual images are generated from center image by pixels shifting in the DIBR process.

In paper [7], they divide the image into several rings with the center of the image as the center, compressing the descriptors of selected SIFT feature points in every ring to generate the final hash. In fact, the center of virtual images and center image are different result in the hash vectors between the center image and virtual images are different. In paper [9], they also 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.

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.

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 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 Fig. 5, Fig. 6, Fig. 7 and Fig. 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 clearly that the maximum distances between the original key-based hash and other 400 hashes with different keys are lower than 0.96 as shown in Fig. 3. This experimental result shows that the security of 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 images identification, authentication 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 ensure that our DIBR 3D image hashing scheme also have better performance for virtual images identification. The experiment results have shown that the proposed DIBR 3D image hashing resists 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.

Acknowledgment. We would like to thank anonymous reviewers for their helpful comments and suggestions, and their comments and suggestions help us to improve this paper's quality. 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, RCCXYJ201812). The Open Fund of the State Key Laboratory of Information Security (2019-ZD-05). The Natural Science Foundation of Zhejiang Province (No. LY18F020020).

References

1. Fehn, C.: 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 (2004)
2. Ahmed, F., Siyal, M.Y., Abbas, V.U.: A secure and robust hash-based scheme for image authentication. Sig. Process. **90**(5), 1456–1470 (2010)
3. Monga, V., Evans, B.L.: Perceptual image hashing via feature points: performance evaluation and tradeoffs. IEEE Trans. Image Process. **15**(11), 3452–3465 (2006)
4. Kozat, S., Venkatesan, R., Mihcak, M.: Robust perceptual image hashing via matrix invariants. In: 2004 International Conference on Image Processing, pp. 3443–3446. IEEE, Singapore (2004)
5. Monga, V., Mhcak, M.K.: Robust and secure image hashing via non-negative matrix factorizations. IEEE Trans. Inf. Forensics Secur. **2**(3), 376–390 (2007)
6. Roy, S., Sun, Q.: Robust hash for detecting and localizing image tampering. In: 2007 IEEE International Conference on Image Processing, pp. 117–120. IEEE, San Antonio (2007)
7. Lv, X., Wang, Z.J.: Perceptual image hashing based on shape contexts and local feature points. IEEE Trans. Inf. Forensics Secur. **7**(3), 1081–1093 (2012)
8. Tang, Z.J., Zhang, X.Q., Li, X.X., Chao, S.C.: Robust image hashing with ring partition and invariant vector distance. IEEE Trans. Inf. Forensics Secur. **11**(1), 200–214 (2016)
9. Tang, Z.J., Zhang, X.Q., Chao, S.C.: Robust perceptual image hashing based on ring partition and NMF. IEEE Trans. Knowl. Data Eng. **26**(3), 711–724 (2014)
10. Zhang, L., Tam, W.: Stereoscopic image generation based on depth images for 3D TV. IEEE Trans. Broadcast. **51**(2), 191–199 (2005)
11. Scharstein, D., Pal, C.: Learning conditional random fields for stereo. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. IEEE, Minneapolis (2007)
12. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2007)
13. Xiang, S.J., Kim, H.J., Huang, J.W.: Invariant image watermarking based on statistical features in the low-frequency domain. IEEE Trans. Circuits Syst. Video Technol. **18**(6), 777–790 (2008)
14. Ground Truth Database. http://www.cs.washington.edu/research/imagedatabase/groundtruth/. Accessed 8 May 2008