

Received April 12, 2020, accepted May 8, 2020, date of publication May 15, 2020, date of current version June 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2994966

NSCT-Based Robust and Perceptual Watermarking for DIBR 3D Images

SEUNG-HUN NAM¹, SEUNG-MIN MUN², WONHYUK AHN¹, DONGKYU KIM³,
IN-JAE YU¹, WOOK-HYUNG KIM⁴, AND HEUNG-KYU LEE¹

¹School of Computing, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, South Korea

²Mobile Business, Samsung Electronics Company, Ltd., Suwon 443742, South Korea

³Samsung Research, Samsung Electronics Company, Ltd., Seoul 06765, South Korea

⁴Visual Display Business, Samsung Electronics Co., Ltd., Suwon 16677, South Korea

Corresponding author: Heung-Kyu Lee (heunglee@kaist.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) grant (NRF-2019R1A2C2084569), and in part by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant (No. 2017-0-01671, Development of high reliability image and video authentication service for smart media environment) funded by the Korea Government (MSIT).

ABSTRACT Depth-image-based rendering (DIBR), where arbitrary views are synthesized from a center image and depth image, has received much attention in the three-dimensional (3D) research field. With advances in depth-acquisition techniques and the proliferation of 3D glasses and 3D display devices, there is a growing demand for schemes to protect the copyrights of DIBR 3D images. Digital watermarking is a typical protection technology and designing a watermarking method for DIBR 3D images is a challenging task because the synchronization of watermarks can easily be broken in the process of generating synthetic images. To address this issue, we propose a non-subsampled contourlet transform (NSCT)-based blind watermarking for DIBR 3D images. To ensure the proposed method has properties of robustness against the DIBR process, we conduct an analysis of robustness for NSCT subbands against DIBR attacks. Based on the analysis results, we select subbands that are robust against DIBR attacks and embed watermark in the low coefficients of NSCT subbands using quantization-based embedding. While ensuring robustness, the proposed method also improves the imperceptibility of watermarks by adjusting their embedding strength according to computed perceptual masking values. Through experiments, we show that the proposed method is robust against both desynchronization attacks of the DIBR process and common attacks including signal processing operations and geometric distortions. The high imperceptibility of our method is also verified by several evaluation metrics in a subjective and objective manner.

INDEX TERMS Depth-image-based rendering (DIBR), digital watermarking, DIBR 3D image watermarking, non-subsampled contourlet transform (NSCT), multi-bit watermarking.

I. INTRODUCTION

Three-dimensional (3D) content has been attracting a lot of attention from various industries and viewers. Due to the immersive experience provided by 3D content and advances in 3D displays, the 3D content market has continued to increase, as has the demand for 3D content. Such 3D content are generated by two methods: stereo image recording (SIR) and depth-image-based rendering (DIBR) [1], [2]. SIR, which is the conventional approach to generating stereoscopic images, simultaneously captures the same scene using two cameras located horizontally. Since the stereoscopic

The associate editor coordinating the review of this manuscript and approving it for publication was Charith Abhayaratne¹.

images captured by the SIR are a real scene, it allows viewers to enjoy a high-quality viewing experience when viewed using a 3D display. However, the cameras' fixed positions means that the parallax of the stereoscopic images is fixed, and so SIR has a disadvantage that it cannot adjust the depth condition to suit the viewer's preference [3].

Meanwhile, DIBR generates synthetic images using a center image, also referred to as a reference image, and its associated depth image [4]–[6]. Since 3D display devices (e.g. 3D-TVs, auto-stereoscopic displays, and free-viewpoint TVs) require both views actually captured from a camera and views from arbitrary virtual viewpoints, DIBR is an evolutionary and optimized approach to a 3D broadcasting system [7], [8]. Since the

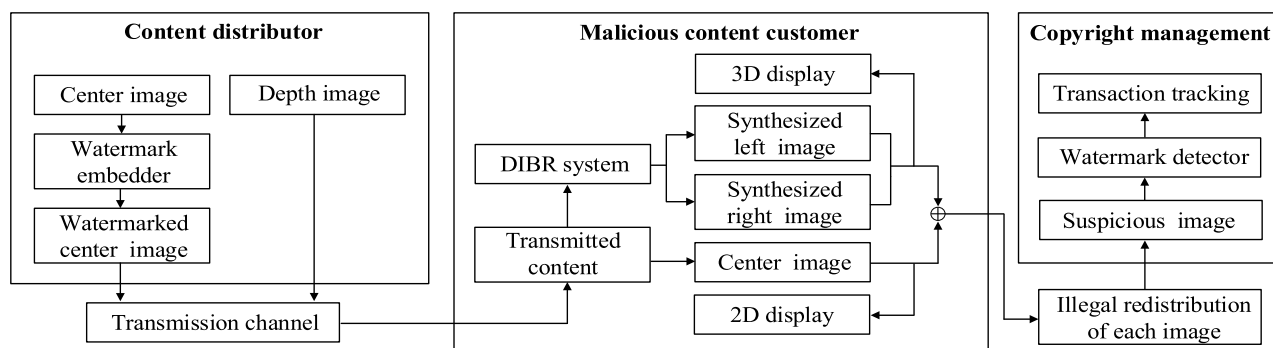


FIGURE 1. Illustration of the DIBR-based 3D broadcasting system and scenario of the illegal dissemination and copyright management of DIBR 3D images.

multiview-video-plus-depth (MVD) format has been specified as a standard of input for efficient compression and transmission, DIBR has become more widely used [9]. Recently, as the technique for acquiring depth images from a single image and the inpainting technique are advanced, various applications based on DIBR have been actively developed [10], [11]. From the perspective of generating stereoscopic images, DIBR generates both left and right images using a color-center image, a grayscale depth image, and depth condition (baseline distance). DIBR has the advantage that it allows viewers to adjust the parallax of two synthesized views to achieve depth perception, taking user preference into consideration [4], [5]. In addition, the DIBR method has lower network bandwidth and storage costs than the SIR method [6].

As interest in DIBR has increased, the issue of copyright protection for DIBR-based synthesized images, referred to as DIBR 3D images, has emerged. In the past, numerous watermarking methods have been proposed to solve the copyright problem of two-dimensional (2D) images. However, 2D image watermarking methods are not directly applicable to DIBR 3D images because they have not been designed to consider inherent desynchronization attacks that can occur in the process of generating synthetic images [1], [2]. Desynchronization attack is known as one of the most difficult to resist in that it can desynchronize the location of a watermark embedded in an image, causing incorrect watermark extraction. The desynchronization attacks that result from DIBR are as follows:

- **3D image warping** Pixels in the center image are moved horizontally to generate a virtual image that results in desynchronization of the watermark embedded in the center image.
- **Preprocessing of depth image** This step is used to reduce the sharp depth discontinuities in the depth image and preprocessed depth information affects the process of generating synthetic images.
- **Baseline distance adjustment** By adjusting the baseline distance, the parallax of the synthesized left and right views can be controlled; This affects the distance that pixels move horizontally.

This implies that robust watermarking is in high demand for DIBR 3D images, which should be robust against the above three attack types, referred to as DIBR attacks in this paper. Making watermarking robust against DIBR requires considering what illegal redistribution may occur in a DIBR-based 3D broadcasting system [2]. As shown in Fig. 1, a content distributor transmits the watermarked center image and its corresponding depth image to a customer via a DIBR-based 3D broadcasting system. On the receiver side, a malicious customer could generate virtual images using the DIBR process and then illegally redistribute not only the center image but also the synthesized left and right images. Thus, watermarking for DIBR 3D images should be able to extract watermarks for the following listed content: the center image and the synthesized left and right images [1].

In this paper, we focus on designing a watermarking method that is robust against desynchronization attacks that occur in the DIBR process, taking into consideration the watermarking methods' fundamental requirements [15], [16] including robustness, imperceptibility, blind extraction, and sufficient watermark capacity. As mentioned above, since the synthetic image generation of DIBR is a partial translation along the horizontal direction considered a desynchronization attack, watermarking for DIBR 3D images should have robustness against DIBR attacks. In our work, we utilize the non-subsampled contourlet transform (NSCT) domain, which is the first attempt in the field of watermarking for DIBR 3D images. We believe that the NSCT domain, which has not previously been used for DIBR 3D image watermarking, is advantageous for securing invisibility and robustness compared to other domains such as discrete cosine transform (DCT) and dual tree complex transform (DTCWT) domains. DCT and DTCWT are domains used in watermarking methods [12], [13] selected as baselines in the DIBR watermarking field, but these two domains are accompanied by degradation of image quality in the watermark embedding process.

As reported in [14], DCT domain-based watermark embedding triggers block artifacts, and in the case of the DTCWT domain, the imperceptibility decreases due to the watermark insertion into the subbands generated through sampling of the

TABLE 1. A comparison of the proposed method and existing methods for the fundamental requirements of DIBR 3D image watermarking.

Method	Fidelity		Blind extraction	Capacity	Robustness				
	Objective	Subjective			3DW	PDM	BDA	SPO	GD
IR-based [12]	○	∞	○	∞	○	▽	▽	○	▽
DTCWT-based [13]	▽	▽	○	○	○	○	○	○	○
SIFT-based [14]	∞	○	○	▽	○	○	○	▽	○
Proposed method	∞	∞	○	○	∞	∞	∞	○	∞

Notes: The abbreviations and symbols used in Table 1 are as follows. 3DW = 3D image warping, PDM = preprocessing of depth map, BDA = baseline distance adjustment, SPO = signal processing operation, GD = geometric distortion; ∞ = good, ○ = common, ▽ = bad.

decomposition process. Compared to the DTCWT domain, the NSCT domain without subsampling in the decomposition process has little degradation in image quality due to watermark insertion. In addition, because the NSCT domain has the property of shift invariance, it was considered suitable for DIBR 3D image watermarking, which was confirmed through the results of the robustness analysis and various experiments in Sections IV and VI. Through a robustness analysis of NSCT subbands against DIBR attack, we select NSCT subbands with robustness against horizontal pixel shift, and then we adaptively quantize the row coefficients of NSCT subbands in the watermark embedding phase.

In the field of watermarking, it is important not only to select a domain for watermarking, but also to design a framework well by combining each module of watermark insertion and extraction with consideration of robustness and imperceptibility. That is, it is possible to achieve both robustness and imperceptibility by carefully selecting modules for the insertion and extraction of watermark through the analysis of the domain and DIBR process, rather than simply utilizing the properties obtainable in the NSCT domain. For robustness, quantization-based watermark embedding is performed for row coefficients, not column coefficients, to be robust against the horizontal pixel shift occurring in DIBR attacks. In the watermark extraction phase, the watermarks are extracted using the statistical difference between the coefficients caused by the watermark embedding phase. With consideration of imperceptibility and robustness against common attacks, including signal processing operations and geometric distortions, the embedding strength of watermarks is controlled by the perceptual masking value and experimentally determined parameters, thereby minimizing the degradation of visual quality caused by the watermark embedding.

As we intended, our method shows higher watermark extraction performance against DIBR attacks, signal processing operations, and geometric distortions while maintaining higher imperceptibility than comparative methods [12]–[14]. In addition, the proposed method can extract embedded watermarks in a blind fashion while ensuring sufficient watermark capacity. The main contributions of the proposed method are as follows.

- Our work is the first study to apply the NSCT domain with the shift invariance property to watermarking for DIBR 3D images, and we showed that the NSCT

domain is suitable for the given task through the results of the robustness analysis and various experiments in Sections IV and VI.

- Our goal was to design a method that meets the fundamental requirements of DIBR 3D image watermarking. To do this, we conducted extensive experiments in the process of applying quantization-based embedding and perceptual masking to NSCT subbands and specifying parameters that yielded the best performance. To verify the effectiveness of the proposed method, extensive experiments were designed based on existing comparative methods [12]–[14] with unique strengths (see Table 1) and larger datasets [4], [17]–[20], and the experimental results showed the superiority of the proposed method in terms of fundamental requirements of DIBR 3D image watermarking.

The remainder of this paper is organized as follows. Section II reviews related works, Section III presents the DIBR system, and Section IV briefly reviews the NSCT domain. The proposed watermarking method is proposed in Section V and the performance of the proposed method is then demonstrated in Section VI. Finally, Section VII concludes this paper.

II. RELATED WORK

Over the past decade, several watermarking methods have been introduced for DIBR 3D images. Depending on whether blind extraction (ability to extract watermark in a given work without extra information) is possible, these methods can be classified as non-blind watermarking, semi-blind watermarking, and blind watermarking [15]. Initially, non-blind watermarking methods were proposed in which the original image is necessary in the watermark extraction process. In [21], Halici *et al.* proposed a watermarking method based on the estimation of projection matrix between center image and synthesized image. In [22], Lee *et al.* proposed perceptual watermarking based on human visual system (HVS) that predicted the occlusion regions that were occluded by adjacent pixels after view synthesis.

In [23]–[25], several local feature descriptor-based semi-blind watermarking methods that necessary side information in the watermark extraction process were proposed. Under the predefined rendering condition, the method in [23] utilized matched common areas between the center image and the

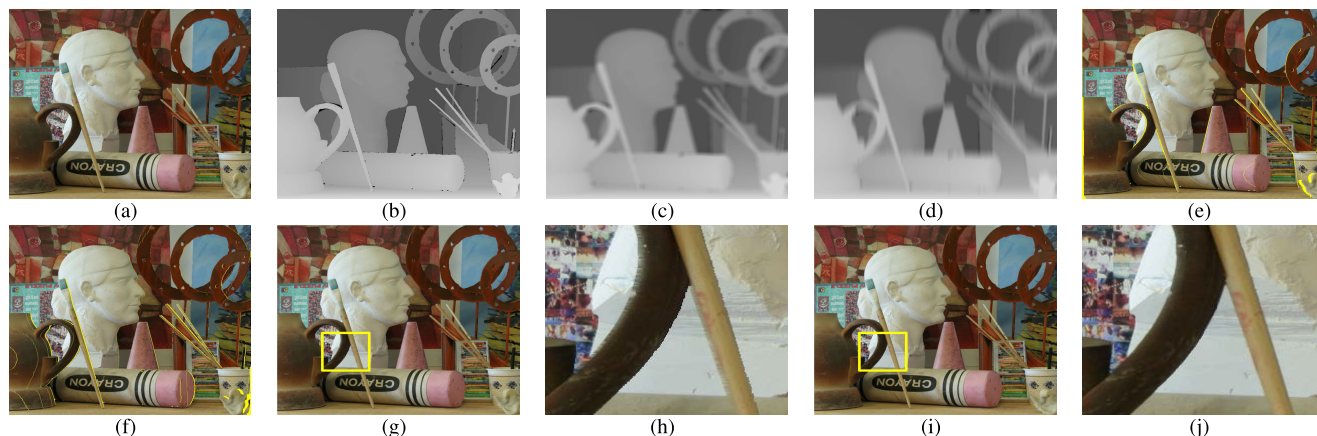


FIGURE 2. Illustration of image synthesis based on DIBR using an Art image: (a) Center image, (b) Depth image, (c) Depth image smoothed by a symmetric Gaussian filter, (d) Depth image smoothed by an asymmetric Gaussian filter, (e) Left image with holes represented in yellow, (f) Right image with holes represented in yellow, (g) Left image with hole-filling, (h) Magnified regions of (g) with a yellow box, (i) Result of applying hole-filling to the left image generated by asymmetric filter-based preprocessing of the depth image, (j) Magnified regions of (i) with a yellow box.

synthesized left and right images based on descriptor matching. In [24], Miao *et al.* proposed a resynchronization scheme that uses descriptors to estimate the disparity map between the center image and a synthesized image; this approach could extract a watermark in any synthesized image with disparity map-based compensation. In [25], Cui *et al.* presented discrete wavelet transform-based watermarking with geometric rectification. Since this method rectified a geometrically distorted view, it had robustness against geometric distortions and desynchronization attacks in the DIBR process.

Unlike the above-mentioned methods [21]–[25] that have limitations in terms of practical applications, watermarking methods considering blind extraction and the fundamental requirements of watermarking have been proposed. In [26], Lee *et al.* proposed horizontal noise mean shifting-based watermarking that exploits the mean of the horizontal noise histogram of the center image. In [27], Rana *et al.* proposed depth map-based dependent region detection and block partitioning. The DCT is applied to blocks of the dependent regions that are robust against view synthesis and then a watermark bit is inserted by modifying the DC coefficients. The authors in [28] proposed a histogram shape-based blind watermarking method that exploited a pixel mean value-based pixel group selection method. In [29], Kim *et al.* presented a template-based approach for robustness against geometric distortions; this approach inserts the template and watermark into the curvelet domain and one-dimensional (1D)-DCT domain, respectively.

In [12], Lin and Wu proposed blind watermarking based on the DCT domain and the inverse rendering (IR) of the DIBR process. Under a predefined rendering condition, three watermarks for the center, left, and right images are embedded in the DCT domain of the center image. In [13], Kim *et al.* presented a DTCWT-based blind watermarking. To design the method such that it is robust against DIBR attacks, the approximate shift invariance of the DTCWT domain is employed. These two methods in [12], [13] are

most actively compared as the baseline. Recently, the authors in [14] proposed a scale-invariant feature transform (SIFT)-based method that utilizes the invariability of the parameters of the SIFT keypoints to deal with the issue of view synthesis. Unlike the approaches in [23]–[25], this method can extract watermarks blindly without side information, but has the obstacle of limited watermark capacity.

To design a robust and perceptual watermarking method for DIBR 3D images that outperforms existing methods, we conducted a performance analysis of the existing comparative methods [12]–[14] that have their strengths in terms of fundamental requirements. Table 1 presents the results of analyzing each watermarking method's strengths and weaknesses. As listed in Table 1, the IR-based method [12] has advantages in terms of watermark capacity and the SIFT-based method [14] has high performance in objective fidelity test and robustness against geometric attacks such as translation and cropping because it inserts watermarks around specific keypoints. The DTCWT-based method [13] performed generally well against various attacks including DIBR attacks, signal processing operations, and geometric distortions. As listed in Table 1, it is difficult to achieve high robustness and invisibility simultaneously due to trade-offs between the fundamental requirements of watermarking. To differentiate from methods [12]–[14], we aim to design a watermarking method that is robust against DIBR attacks and common attacks while also maintaining high imperceptibility. In addition, the proposed method is designed to extract watermarks directly from a given image without additional information and ensure sufficient watermark capacity.

III. DEPTH-IMAGE-BASED RENDERING

Depth-image-based rendering (DIBR) generates synthesized images captured from virtual viewpoint using a center image and its associated depth image [4], [5]. Fig. 2(a) and Fig. 2(b) show examples of a color center image and associated grayscale depth image, respectively. In the depth image,

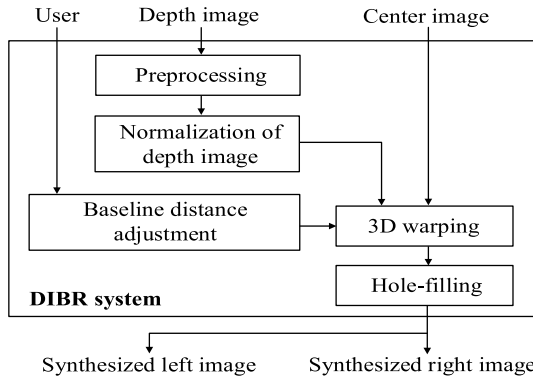


FIGURE 3. Block diagram of the DIBR system for generating synthesized images.

a higher intensity value denotes that objects are closer to the shooting camera. As illustrated in Fig. 3, the DIBR system consists of the following steps: preprocessing of depth image, depth normalization and 3D image warping, and hole-filling [30]. For natural synthesized image generation, preprocessing the depth image is needed before virtual view rendering [3]. 3D image warping is employed to generate synthesized images by partially moving some pixels of the center image horizontally, and baseline distance adjustment is employed to control the rendering condition [4], [5]. The goal of hole-filling is to fill in holes, referred to as disocclusions, generated from the 3D image warping [31].

A. PREPROCESSING OF DEPTH IMAGE

Preprocessing of the depth image is employed to reduce hole occurrences in view synthesis [6]. At this stage, the depth image is usually smoothed by a Gaussian filter to mitigate sharp depth discontinuities at objects’ edges and borders in the depth image [3]. A 1D Gaussian filter is defined as follows: $g(x, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp(\frac{-x^2}{\sigma^2})$, where $-\frac{w}{2} \leq x \leq \frac{w}{2}$. Here, σ and w are the standard deviation and window size of the filter, respectively. Let $D(x, y)$ be a depth value in the depth image at pixel coordinates (x, y) and $\tilde{D}(x, y)$ in the ranges 0-255. As depicted in [6], the depth value $\tilde{D}(x, y)$ in the preprocessed depth image after Gaussian filter-based smoothing equals

$$\frac{\sum_{v=-\frac{w}{2}}^{\frac{w}{2}} \left\{ \sum_{h=-\frac{w}{2}}^{\frac{w}{2}} (D(x-h, y-v)G(h, \sigma_h))G(v, \sigma_v) \right\}}{\sum_{v=-\frac{w}{2}}^{\frac{w}{2}} \left\{ \sum_{h=-\frac{w}{2}}^{\frac{w}{2}} G(h, \sigma_h)G(v, \sigma_v) \right\}}, \tag{1}$$

where $G(h, \sigma_h)$ and $G(v, \sigma_v)$ are the Gaussian filter for the horizontal and vertical directions. Here, σ_h and σ_v denote the horizontal and vertical standard deviations, respectively. Symmetric filter-based preprocessing is presented in [32] and Fig. 2(c) shows the depth image after smoothing with a symmetric Gaussian filter where $\sigma_h = \sigma_v = 20$. The authors in [3], [6] showed that symmetric smoothing can generate

a distortion leading to the vertical boundaries becoming curved; they proposed asymmetric filter-based preprocessing of the depth image to generate more natural views. Fig. 2(d) indicates the depth image after smoothing with an asymmetric Gaussian filter with $\sigma_h = 10$ and $\sigma_v = 30$. The effectiveness of preprocessing the depth image is addressed in the last subsection.

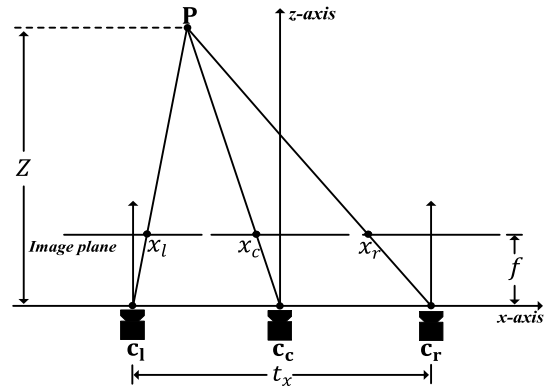


FIGURE 4. Parallel camera configuration for the generation of virtual left and right images. One point P with depth Z is projected onto the image plane of each of the three cameras.

B. 3D IMAGE WARPING AND DEPTH NORMALIZATION

3D image warping is the process of generating virtual views from a center image and its corresponding per-pixel depth information. This process consists of the two following steps: 1) With depth values, the points in the center image are re-projected into 3D space, 2) These points on 3D space are projected into the image plane of the virtual left and right cameras [4]. The parallel camera configuration is typically utilized for view synthesis because it does not generate vertical disparities, unlike the convergent camera configuration [30]. Fig. 4 represents the illustration of 3D image generation on a parallel camera configuration, where c_l , c_c , and c_r denote the left, center, and right cameras, respectively. Before view generation, the depth value \tilde{D} of the preprocessed depth image is normalized linearly such that the depth values lie within the new range from the farthest clipping plane Z_f to the nearest clipping plane Z_n [1], [4]. Normalization is performed according to the following equation:

$$Z = Z_f - \tilde{D} \frac{(Z_f - Z_n)}{255}, \tag{2}$$

where Z denotes the normalized depth value.

In 3D image warping, pixels in the center image are horizontally moved according to the corresponding normalized depth value. Under a parallel camera configuration, the y-coordinate of the projection of the point P with depth Z on each image plane is identical. Based on this camera configuration approach, virtual views can be generated from the following function [3], [4]:

$$x_l = x_c + \frac{t_x f}{2 Z}, \quad x_r = x_c - \frac{t_x f}{2 Z}, \tag{3}$$

where t_x and f denote the baseline distance between two virtual cameras \mathbf{c}_l and \mathbf{c}_r and these cameras' respective focal length; t_x is used to control the depth condition. x_l , x_c and x_r denote the pixel x -coordinate of the virtual left view, center view and virtual right view, respectively. Figs. 2(e) and 2(f) are the synthesized left and right images, respectively; synthesized images are generated with baseline distance $t_x = 5\%$ of the image width.

C. HOLE-FILLING

The hole-filling process is employed to fill-in newly exposed areas, referred to as holes, that appear in virtual views [31]. The yellow pixels of Fig. 2(e) and 2(f) represent the holes that occur due to 3D image warping. One solution is to replace the holes by linear interpolation with adjacent pixels. Here, we exploit linear interpolation as a hole-filling method due to its efficiency. Compared to Fig. 2(e), we can see that the quality of Fig. 2(g) with interpolation-based hole-filling has improved. Furthermore, the effectiveness of the depth image preprocessing is verified by comparing with two magnified figures Fig. 2(h) and 2(j). The perceptible distortions in synthesized images are mitigated by utilizing depth image smoothing and hole-filling.

IV. NON-SAMPLED CONTOURLET TRANSFORM

The non-sampled contourlet transform (NSCT) domain has shift invariance, multi-scale, and multi-directional selectivity properties, hence it has robustness against geometric distortion-like DIBR attacks including 3D image warping, depth image preprocessing, and baseline distance adjustment. Here, we present a brief review of NSCT and analysis of the NSCT subbands that are particularly robust to desynchronization attacks resulting from DIBR.

A. BRIEF REVIEW OF NSCT

Contourlet transform (CT) is a multi-scale and multi-directional transform that is constructed by the Laplacian pyramid and directional filter bank (DFB) [33]. The CT performs well in image processing applications but this transform is not shift-invariant since downsampling and upsampling are needed in both DFB and the Laplacian pyramid. To overcome the drawback of the CT, the authors in [34] proposed NSCT, which is a shift-invariant version of CT. NSCT is established as a combination of two structures of a non-sampled pyramid (NSP) and non-sampled directional filter bank (NSDFB). The left part of Fig. 5 illustrates the NSP-based multi-scale decomposition. The NSP is a two-channel non-sampled filter with ensured multi-scale property. At each NSP decomposition level, one low-pass subband and one high-pass subband are generated that have the same input size [35]. Given that the decomposition level is l_d , NSP generates $l_d + 1$ subbands and Fig. 5 gives an example of multi-scale decomposition with $l_d = 3$ (one low-pass subband and three high-pass subbands).

The directional decomposition is derived from the NSDFB and this filter bank decomposes the high-pass subband

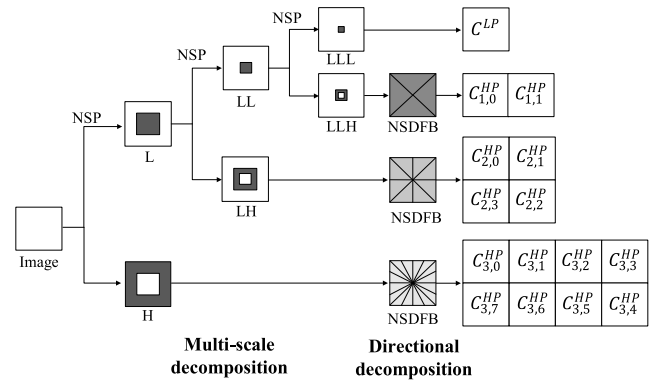


FIGURE 5. NSCT implementation with the NSP and NSDFB structures.

into several directional high-pass subbands. NSDFB is constructed based on a two-channel fan filter bank and resampling; this filter bank splits the 2D frequency plane into directional wedges [35]. NSDFB is iteratively exploited on the high-pass subband to provide directional decomposition [34]. As shown in the right part of Fig. 5, each high-pass subband generated from NSP-based multi-scale decomposition is filtered by NSDFB to construct 2^1 , 2^2 , and 2^3 directional subbands from coarse scale 1 to fine scale 3. In this work, NSCT decomposition level l_d is set to 3 and the coefficients of the low-pass subband and directional high-pass subband in the NSCT domain are represented by C^{LP} and $C_{s,d}^{HP}$, respectively. Here, s and d denote scale and direction index, respectively.

Based on the properties of the CT and NSCT domains, watermarking methods for 2D images have been steadily presented [36]–[41]. In particular, NSCT-based watermarking methods for 2D medical images in [36], [37] showed outstanding performance by exploiting NSCT and DCT in a hybrid method. Unlike existing NSCT-based approaches for 2D images in [36]–[39], we propose DIBR 3D image watermarking exploiting the NSCT domain. As mentioned in the previous sections, the DIBR process can be regarded as a desynchronization attack in that it affects the distance of pixels moving in the horizontal direction for view synthesis. Since the approaches in [36]–[39] are not designed for DIBR attacks, such techniques cannot be applied directly to DIBR 3D images. Therefore, in order to design a watermarking method for DIBR 3D images, domains that are robust to DIBR attacks must be analyzed, followed by research on watermark insertion and extraction algorithms. The following subsection provides an analysis of the robustness of the directional high-pass subbands against DIBR attacks.

B. ANALYSIS OF ROBUSTNESS FOR NSCT SUBBANDS AGAINST DIBR ATTACKS

NSCT is a shift-invariant transform with multi-scale and multi-directional selectivity based on NSP and NSDFB structures [42]. As we saw in Section III, DIBR attacks can be regarded as a partial translation for some pixels in the center image that causes watermark desynchronization [14].

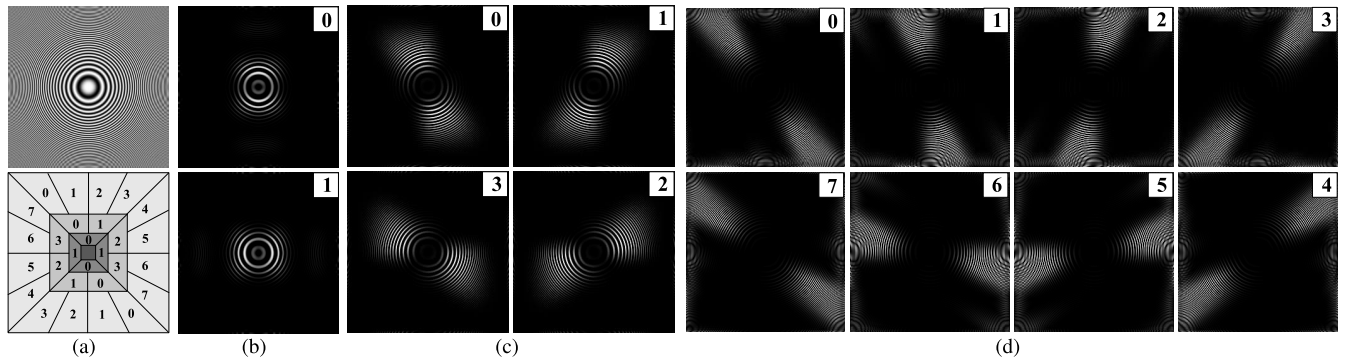


FIGURE 6. NSCT decomposition on *Zoneplate* image (256 × 256): (a) Original image (top) and order of directional wedges for scale 1, 2, and 3 (bottom), (b) The high-pass subbands with two-directional decomposition for scale 1, (c) The high-pass subbands with four-directional decomposition for scale 2, (d) The high-pass subbands with eight-directional decomposition for scale 3.

Hence, the watermarking scheme for DIBR 3D images should be designed to be robust to watermark desynchronization that results from DIBR attacks. The NSCT domain has the advantage of being more resistant to DIBR attacks than other transform domains due to the shift invariance property. With multi-directional property, NSCT decomposition also provides precise directional details for a given image [42]. NSCT high-pass subbands can effectively capture edges with the shape of smooth contours in various directions that are the dominant features of a given image.

The representations of directional subbands depicted in Fig. 6 show the edge details for various directions. The top and bottom of Fig. 6(a) show the original image and order of directional wedges for each scale, respectively. Fig. 6(b)-Fig. 6(d) represent directional high-pass subbands for scales 1, 2, and 3, respectively. The upper right number of each subband representation indicates a direction index of the corresponding subband. Looking at Fig. 6(c) as a representative, NSCT decomposition for scale 2 has four directional high-pass subbands from $C_{2,0}^{HP}$ to $C_{2,3}^{HP}$. Each subband represents the characteristics of the *Zoneplate* image around the edges along with its directional wedge of the filter; evidently, each subband has strong energy in a specific direction. This means that each directional high-pass subband may have a different robustness against DIBR attacks.

To determine what high-pass subbands in the NSCT domain are suitable for watermarking and robust against DIBR attacks, similarity measurement experiments between the high-pass subbands obtained from the center and synthesized images were conducted. If the similarity of a specific subband is measured as high, this indicates that this subband is resistant to the partial translation of pixels on the DIBR process. In this experiment, we employed center and depth images from the following datasets [4], [17]–[20]: Microsoft Research 3D Video Datasets, Heinrich-Hertz Institut, and Middlebury Stereo Datasets. In the process of generating synthesized left and right images using DIBR, the focal length f is set to 1 and the value of t_x to 2%–5% of the width of the center image to vary the depth condition, and synthesized

images are generated for each t_x . As a metric for measuring the similarity between subband coefficients, the mean square error (MSE) was employed and the MSE between $C_{s,d}^{HP}$ and \hat{C}_{s,d,t_x}^{HP} is equal to

$$\frac{1}{W \times H} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} \|C_{s,d}^{HP}(i, j) - \hat{C}_{s,d,t_x}^{HP}(i, j)\|^2, \quad (4)$$

where $C_{s,d}^{HP}$ and \hat{C}_{s,d,t_x}^{HP} indicate the high-pass subband decomposed from the original center image and high-pass subband decomposed from the DIBR-based synthesized image with t_x , respectively. W and H denote the width and height of the given image, respectively, and i and j denote the coordinates of the subband coefficients. Lower MSE values indicate a lesser change in the coefficients of the subbands in the NSCT domain after the DIBR process.

Fig. 7 shows the results of the average MSE between the subband coefficients of the center and synthesized images. Each legend shown in the upper-right-hand corner of Fig. 7 represents t_x used in the process of generating DIBR-based synthesized images. First, we see that MSE increases as the value of t_x increases. This is because the DIBR attacks applied to the center image become stronger as

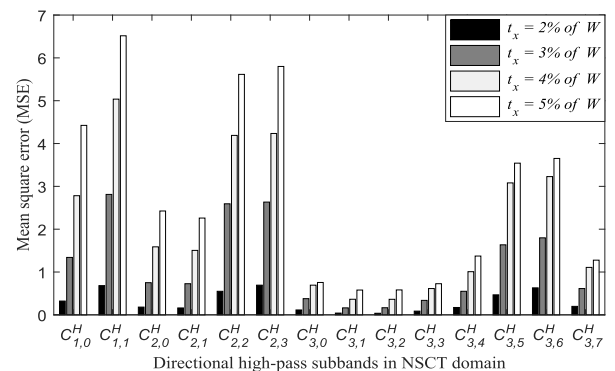


FIGURE 7. Average MSE between the subband coefficients in the NSCT domain of the center image and the synthesized image where W denotes the width of the center image and the synthesized images.

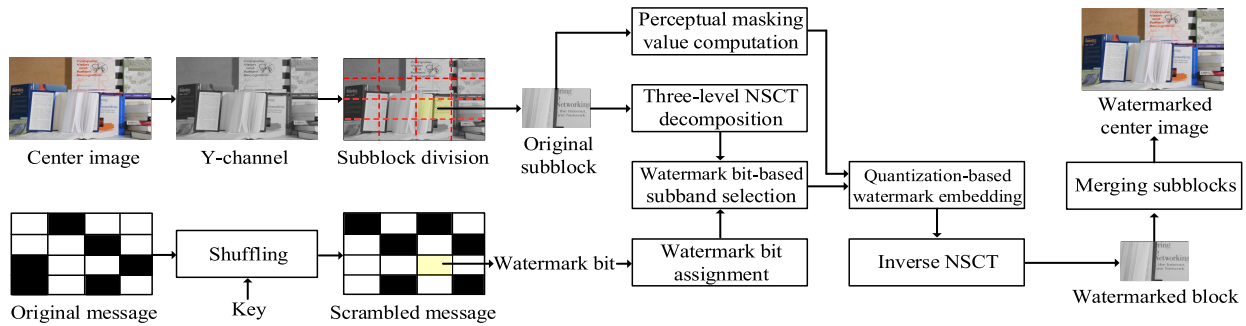


FIGURE 8. Block diagram of the watermark embedding process in the proposed NSCT-based robust and perceptual watermarking for DIBR 3D images.

the t_x value increases. In addition, we confirmed that the subbands for scale 1 had higher MSE values than those of other scales, hence the subbands belonging to scales 2 and 3 are suitable for DIBR watermarking. In our work, the subbands for scale 2 are used for robustness against common attacks and we used half of the subbands of scales 2 and 3 for watermarking. For scales 2 and 3, the high-pass subbands showing greater robustness against DIBR attacks than other subbands in the same scale were as follows: $C_{2,0}^{HP}$, $C_{2,1}^{HP}$, $C_{3,0}^{HP}$, $C_{3,1}^{HP}$, $C_{3,2}^{HP}$, and $C_{3,3}^{HP}$.

These six subbands show much lower values of MSE than other subbands on the same scale, indicating that the subbands show little change in coefficient value even if a partial translation of pixels in the horizontal direction of the DIBR process is applied. As we clearly see through Fig. 6, each directional high-pass subband has more energy around the edges along with its directional filter bank. As mentioned previously, since the pixels of the center image move horizontally according to corresponding depth information in the DIBR process, the vertical edges are more degraded or distorted than the horizontal edges. Thus, directional high-pass subbands that have more energy along vertical edges than horizontal edges are relatively more vulnerable to DIBR attacks. Looking at the aspect of the selected six subbands in Fig. 6 reveals that they have more energy around the horizontal edges, and through this, it can be deduced why these six subbands have lower MSE values than other subbands even after the DIBR process is applied. From this, we can conclude that the six subbands are suitable for the watermarking of DIBR 3D images, and the algorithm using the selected subbands is introduced in the following section.

V. PROPOSED WATERMARKING METHOD

This section describes the proposed NSCT-based watermarking for DIBR 3D images. Our goal is to design a watermarking method that meets the following fundamental requirements: blind extraction, sufficient watermark capacity, robustness, and imperceptibility. The proposed method is a multi-bit watermarking in that watermarks composed of multi-bit messages can be inserted and extracted and it can extract embedded watermarks in a blind fashion. To ensure

robustness against DIBR attacks, robust subbands analyzed in Section IV are employed for watermarking. For the acquisition of additional robustness, we insert watermarks into the coefficients of the NSCT subbands by quantizing each row. In addition, adjusting the strength of watermark embedding based on perceptual masking is utilized to improve imperceptibility without compromising robustness. First, we introduce the process of watermark embedding and then describe the process of watermark extraction based on statistical differences between the coefficients of subbands caused by the quantization-based watermark embedding.

A. WATERMARK EMBEDDING

Fig. 8 illustrates the overall process of watermark embedding, which consists of the following steps.

- 1) Y-channel acquisition and subblock division
- 2) Shuffling the message and watermark bit assignment
- 3) Conduction of three-level NSCT decomposition
- 4) Selection of high-pass subbands based on the assigned watermark bit
- 5) Perceptual masking value computation
- 6) Quantization-based watermark embedding
- 7) Conduction of inverse NSCT
- 8) Repetition of steps 3-7 for the entire subblock
- 9) Merging subblocks and watermarked image generation

The proposed method only embeds multi-bit watermarks in the center image I of size $W \times H$, not in the corresponding depth image. YUV representation can be obtained from an RGB representation using the following equation:

$$\begin{bmatrix} Y_c \\ U_c \\ V_c \end{bmatrix} = \begin{bmatrix} 0 \\ 127 \\ 127 \end{bmatrix} + \begin{bmatrix} 0.2989 & 0.5866 & 0.1145 \\ -0.1688 & -0.3312 & 0.5000 \\ 0.5000 & -0.4184 & -0.0816 \end{bmatrix} \begin{bmatrix} R_c \\ G_c \\ B_c \end{bmatrix}, \quad (5)$$

where R_c , G_c , and B_c denote each channel of RGB representation and Y_c , U_c , and V_c denote each channel of YUV representation. In the field of general watermarking research, the U- and V-channels (chrominance channels) are used to improve the visual quality considering HVS, while the Y-channel (luminance channel) is employed for high

robustness [43], [44]. In our method, watermarks are inserted into the luminance channel of the center image, represented by Y_c , for better robustness. The proposed method has high imperceptibility by utilizing perceptual masking-based watermark embedding and only NSCT high-pass subbands which are less noticeable to HVS. The detailed description of achieving high imperceptibility is introduced later in this section.

To give the proposed method sufficient watermark capacity, Y_c is divided into $N \times M$ subblocks, where each subblock is $\lfloor \frac{W}{N} \rfloor \times \lfloor \frac{H}{M} \rfloor$ in size. Let $B_{u,v}$ indicates a divided subblock of Y_c , where $u \in \{0, 1, \dots, N-1\}$ and $v \in \{0, 1, \dots, M-1\}$. Since the proposed method allocates a watermark bit (0 or 1) to each subblock, our approach has watermark capacity $N \times M$ bits. Before assigning a watermark bit to each subblock, the original message \mathbf{M}_0 of size $N \times M$ is shuffled to improve security using secret key (see the lower left in Fig. 8) and our work uses a Knuth shuffle. The shuffled message has same size of \mathbf{M}_0 , and a watermark bit $b_{u,v}$ located at the (u, v) coordinate of the shuffled message is allocated to $B_{u,v}$. As shown in Fig. 8, the watermark bit represented by yellow shade is assigned to the corresponding subblock represented by yellow shading. On the watermark embedding process, $b_{u,v}$ assigned to $B_{u,v}$ is used to select the specific group of NSCT high-pass subbands. For the sake of clarity, B and b denote the subblock and watermark bit with arbitrary u and v values from now on.

As mentioned previously, we exploit the NSCT domain with the property of shift invariance to make the proposed method robust to watermark desynchronization caused by DIBR attacks. Here, three-level NSCT decomposition is applied to each B . As a result of the NSCT decomposition of B , one low-pass subband C^{LP} and 14 directional high-pass subbands $C_{s,d}^{HP}$ are generated, where s and d denote the scale and direction index, respectively. When the watermark signal is inserted into the coefficients of C^{LP} , it gives the watermarking method high robustness while the visual quality of watermarked image is greatly degraded. Hence, NSCT high-pass subbands are only utilized for watermarking to improve imperceptibility. Based on the robustness analysis of high-pass subbands against DIBR attacks, we carefully selected six high-pass subbands that have better robustness compared to other high-pass subbands in the watermark embedding and extraction process. As mentioned in Section IV, high-pass subbands for scale 1 (i.e. $C_{1,0}^{HP}$ and $C_{1,1}^{HP}$) showed lower robustness against DIBR attacks than other high-pass subbands where $s = 2, 3$; hence, we used the high-pass subbands corresponding to scales 2 and 3 for watermarking. The details on selected subbands for watermarking are as follows:

- For $s = 2$, high-pass subbands where $d = 0, 1$ (i.e. $C_{2,0}^{HP}$, $C_{2,1}^{HP}$) were determined as an embedding domain for watermarking.
- For $s = 3$, high-pass subbands where $d = 0, 1, 2, 3$ (i.e. $C_{3,0}^{HP}$, $C_{3,1}^{HP}$, $C_{3,2}^{HP}$, $C_{3,3}^{HP}$) were determined as an embedding domain for watermarking.

Next, the six selected high-pass subbands are grouped into two groups, denoted \mathbf{G}_0 and \mathbf{G}_1 , taking into account their directional characteristics (see the energy representation of the specific direction for each subband depicted in Fig. 6) and robustness against DIBR attacks. The high-pass subbands corresponding to \mathbf{G}_0 and \mathbf{G}_1 are as follows: $\mathbf{G}_0 = \{C_{2,0}^{HP}, C_{3,0}^{HP}, C_{3,1}^{HP}\}$, $\mathbf{G}_1 = \{C_{2,1}^{HP}, C_{3,2}^{HP}, C_{3,3}^{HP}\}$. We embed the watermarks into just the subbands of one group according to the assigned watermark bit b . In other words, if the value of a given b is 0, watermark embedding is performed in the three high-pass subbands that correspond to \mathbf{G}_0 . If the value of a given b is 1, watermarking is applied to the three high-pass subbands that correspond to \mathbf{G}_1 . These can be expressed as follows:

- If the assigned watermark bit b is 0,

$$\begin{cases} \mathbf{G}_0 \leftarrow WE(\mathbf{G}_0) \\ \mathbf{G}_1 \leftarrow \mathbf{G}_1 \end{cases} \quad (6)$$

- If the assigned watermark bit b is 1,

$$\begin{cases} \mathbf{G}_0 \leftarrow \mathbf{G}_0 \\ \mathbf{G}_1 \leftarrow WE(\mathbf{G}_1) \end{cases} \quad (7)$$

where $WE(\cdot)$ denotes the proposed quantization-based watermark embedding applied to high-pass subbands that correspond to the group. As depicted in the two equations above, depending on the value of a given b , high-pass subbands belonging to one group are watermarked and the remaining groups retain their original coefficients.

Prior to the detailed description of the quantization-based watermark embedding, a perceptual masking value-based adjustment of the embedding strength that can improve the imperceptibility is described. The embedded watermark should not be noticeable to the human eye and not degrade the original image's perceptual quality [14]. Perceptual masking makes it possible to embed a watermark with an appropriate strength with the consideration of HVS and perceptual quality; hence, we control the embedding strength of watermark using perceptual masking value. The proposed method utilizes two perceptual masking schemes that are designed to consider the texture and brightness characteristics inherent in given content. Then, We integrate the two computed values to generate a combined perceptual masking value λ_{PM_B} for a given B . On watermark embedding, the embedding strength of a watermark for B is adjusted using the computed value of λ_{PM_B} . In our study, if the value of λ_{PM_B} for B is greater than a predefined threshold, it means that even if the watermark is inserted with strong intensity, given that B has the property that the artifacts created by watermarking are less noticeable to the human eye. Therefore, the watermark is inserted into selected high-pass subbands of B as in the above case with a strong intensity and it is inserted with an appropriate intensity in the opposite case.

In detail, we first adopt noise visibility function (NVF)-based perceptual masking in [45]–[47] to analyze the textured

area of the given B . The NVF is exploited to identify high-frequency information such as textured areas that are visually less perceptible. In other words, NVF focuses on the fact that the human eye cannot easily recognize a noise signal in edge regions; hence, we strongly insert a watermark into particular identified regions based on NVF. As depicted in [14], [47], the NVF values of B are calculated with the following equation:

$$NVF(i, j) = \frac{1}{1 + \tau \sigma_x^2(i, j)}, \quad \tau = \frac{\xi}{\sigma_{x_{max}}^2}, \quad (8)$$

where $\sigma_x^2(i, j)$ indicates the local variance in a kernel (size 3×3) centered on the pixel with coordinates (i, j) and $\sigma_{x_{max}}^2$ denotes the maximum local variance of given B . Here, $i \in \{0, 1, \dots, \lfloor \frac{W}{N} \rfloor - 1\}$ and $j \in \{0, 1, \dots, \lfloor \frac{H}{M} \rfloor - 1\}$. ξ is an empirically determined scaling constant, where ξ is set to 150 and the obtained NVF values are scaled so that they fit in the range 0-1.

Fig. 9 shows the visualization results of the employed representations to calculate the perceptual masking value and we can find that the NVF values are close to 0 and 1 in textured and flat regions, respectively (see Fig. 9(c); the results are scaled at 0-255 for visualization). To control the embedding strength of the watermark as block-wise rather than pixel-wise, we calculate the average of the computed NVF values of given B . The average NVF value of B , represented by NVF_B , is calculated as follows:

$$NVF_B = \frac{N \times M}{W \times H} \sum_{j=0}^{\lfloor \frac{H}{M} \rfloor - 1} \sum_{i=0}^{\lfloor \frac{W}{N} \rfloor - 1} NVF(i, j). \text{ The NVF-based perceptual masking value } \lambda_{NVF_B} \text{ for adjusting the embedding strength is computed as follows:}$$

$$\lambda_{NVF_B} = \alpha \cdot (1 - NVF_B) + \beta \cdot NVF_B, \quad (9)$$

where α and β are parameters for adjusting the embedding strength according to the value of NVF_B . Based on the computed λ_{NVF_B} , the proposed method can adjust the level of the embedding strength to take the texture properties inherent in a given B into consideration.

In addition, we adopt brightness map (BM)-based perceptual masking. According to [43], [44], the human eye is less perceptive of modifications (e.g. watermark embedding in our work) that occurs in very bright and dark areas; however, when the brightness values of a specific area is in a middle grayscale, HVS is more sensitive to watermark embedding. Considering this, the proposed method increases the embedding strength of watermarks when the brightness of a given B is either very high or very low. For this, we generate BM using B and C^{LP} to analyze the brightness property inherent in the content of B . Since B is the Y-channel representation of the original data illustrated in Fig. 9(a), it itself represents the luminance information, which helps in brightness analysis. In addition, C^{LP} is used to generate BM because it denotes the low-pass subband obtained by applying NSCT decomposition to B . As shown in Figs. 9(b) and 9(d), the visualization of B and C^{LP} have similar properties, hence we combine the two representations for the following comprehensive brightness analysis:

$$BM = \theta \cdot B + (1 - \theta) \cdot C^{LP}, \quad (10)$$

where θ denotes the compositing factor for generating BM and the values of the generated map are scaled from 0 to 1.

To design the proposed method to embed the watermark block-wise based on the brightness characteristics of B , we calculate the average value of the obtained map as follows:

$$BM_B = \frac{N \times M}{W \times H} \sum_{j=0}^{\lfloor \frac{H}{M} \rfloor - 1} \sum_{i=0}^{\lfloor \frac{W}{N} \rfloor - 1} BM(i, j). \text{ BM-based perceptual masking } \lambda_{BM_B} \text{ to control the embedding strength with consideration of the brightness sensitivity is obtained with the following equation:}$$

$$\lambda_{BM_B} = \begin{cases} \eta_1 + \frac{\eta_2(BM_B - \epsilon_1)^2}{\epsilon_1^2}, & \text{if } BM_B \leq \epsilon_1 \\ \eta_1, & \text{if } \epsilon_1 < BM_B \leq \epsilon_2 \\ \eta_1 + \frac{\eta_2(BM_B - \epsilon_2)^2}{(1 - \epsilon_2)^2}, & \text{if } \epsilon_2 < BM_B, \end{cases} \quad (11)$$

where ϵ_1 and ϵ_2 denote the brightness control thresholds, and the proposed method can adjust the embedding strength when the computed brightness is either very high or low with these two thresholds. η_1 and η_2 indicate the base constant factor and scaling factor, respectively. Finally, we jointly combine the obtained λ_{NVF_B} and λ_{BM_B} , and then use the combined perceptual masking value λ_{PM_B} to adjust the degree of quantization on watermark embedding with consideration of perceptual masking for the block-wise texture and brightness sensitivity.

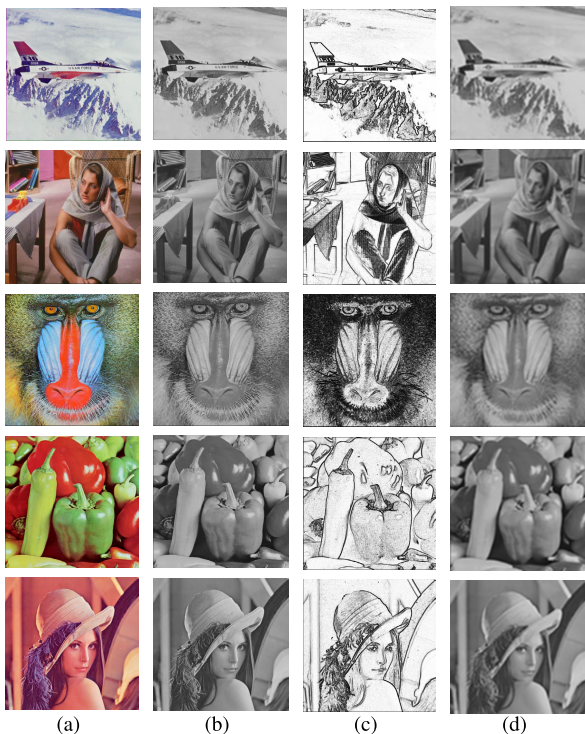


FIGURE 9. Visualization results of representations used to calculate the perceptual masking value: (a) original images in RGB representation, (b) Y-channel representation of (a), (c) visualization of NVF values computed from (b), (d) NSCT low-pass subband computed from (b). From the top row to the bottom row: *Airplane*, *Barbara*, *Baboon*, *Peppers*, and *Lena*.

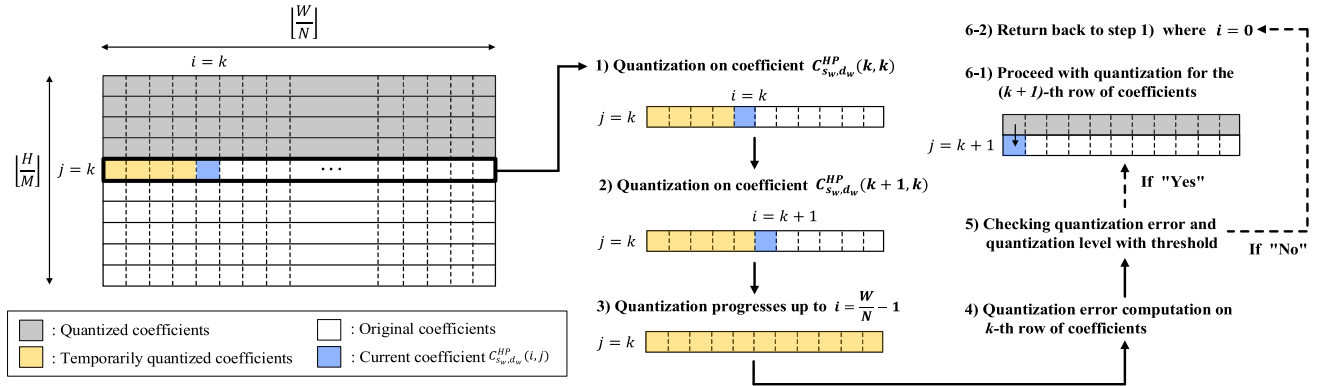


FIGURE 10. Detailed description of quantization-based watermark embedding. In our work, we quantize the rows of coefficients on selected NSCT high-pass subbands to improve the robustness against DIBR attacks.

Designed while considering the characteristics of HVS, λ_{PM_B} can be computed as follows:

$$\lambda_{PM_B} = \lambda_{NVF_B} \cdot \lambda_{BM_B}. \quad (12)$$

The use of λ_{PM_B} to control the strength of the watermark embedding applied to the coefficients of high-pass subbands is described later in this section.

To summarize the contents so far, the proposed watermarking method is based on perceptual masking value, an assigned watermark bit, and six selected high-pass subbands of the NSCT domain as shown in middle part of Fig. 8. The proposed method proceeds with three-level NSCT decomposition and then inserts a watermark into the coefficients of the high-pass subbands that correspond to \mathbf{G}_0 or \mathbf{G}_1 according to the value of allocated b for B . In this step, our method employs NVF- and BM-based perceptual masking values to analyze the inherent visual characteristics of B , and then compute combined perceptual masking value λ_{PM_B} for adjusting the embedding strength of the watermark in consideration of HVS. Next, a methodology of inserting a watermark bit into the coefficient of high-pass subbands corresponding to the group determined by the watermark bit b , which is the core content of the proposed method, is introduced. To design the proposed method to be robust against various attacks, we adopted a quantization-based watermark embedding approach as presented in [13], [48], [49].

As analyzed in Section III, the DIBR system moves pixels in the center image horizontally according to the corresponding depth image to generate a new synthesized view. This process causes fewer statistical changes in the rows of the coefficients than in the columns of coefficients on NSCT subbands [13]. In other words, if watermarks are inserted and extracted in units of columns, it is not robust to a DIBR attack that causes a change in the horizontal direction. Therefore, the proposed method embeds and extracts watermarks using rows of the coefficients of selected subbands. Fig. 10 illustrates the detailed procedure for quantization-based watermark embedding applied to specific high-pass subband. In detail, if the value of b assigned to the B is 0, the quantization procedure

is applied to all high-pass subbands belonging to \mathbf{G}_0 . In this case, the subbands corresponding to \mathbf{G}_1 are kept at their original values without quantization.

From now on, an arbitrary selected high-pass subband to be watermarked is represented by C_{s_w, d_w}^{HP} , and the watermarked high-pass subband is represented by $\tilde{C}_{s_w, d_w}^{HP}$, where s_w and d_w denote the scale and direction index of any high-pass subband that corresponds to \mathbf{G}_0 and \mathbf{G}_1 , respectively. Here, $C_{s_w, d_w}^{HP}(i, j)$ denotes a coefficient value of the (i, j) coordinate where $i \in \{0, 1, \dots, \lfloor \frac{W}{N} \rfloor - 1\}$ and $j \in \{0, 1, \dots, \lfloor \frac{H}{M} \rfloor - 1\}$. As you can see in the part depicted by the bold line in Fig. 10, quantization-based watermark embedding is performed on the coefficients that correspond to the k -th row of the subband where $j = k$. The white, gray, and yellow areas indicate the original coefficients, quantized coefficients, and temporarily quantized coefficients, respectively, and the area in blue denotes the coefficient currently being quantized. Fig. 10 shows the state where the coefficients of i from 0 to $k - 1$ are temporarily quantized and the coefficient, where $i = j = k$, is currently being quantized.

When the quantization of the current coefficient $C_{s_w, d_w}^{HP}(k, k)$ is temporarily completed, the value of i is increased by 1 to quantize the next coefficient $C_{s_w, d_w}^{HP}(k + 1, k)$ (see the middle part of Fig. 10). If this process proceeds until i becomes $\lfloor \frac{W}{N} \rfloor - 1$, the temporary quantization of the coefficients that correspond to the k -th row is completed. Here, by checking the quantization error of the k -th row of coefficients and the quantization level, we can examine whether the quantization-based watermark embedding, in which sufficient robustness and invisibility are secured, has proceeded. If the experimentally determined threshold is satisfied, the value of j is increased by 1 to proceed with the quantization of coefficients corresponding to the $(k + 1)$ -th row. In contrast, if the condition is not satisfied, the quantization step size is increased by increasing the quantization level and the quantization proceeds again from the coefficient where $i = 0, j = k$. If the procedure depicted in Fig. 10 proceeds for the entire row of coefficients, we can obtain a watermarked high-pass subband $\tilde{C}_{s_w, d_w}^{HP}$.

Algorithm 1 Quantization-Based Watermark Embedding on a Row of Coefficients in a High-Pass Subband**Input:** A original high-pass subband C_{s_w, d_w}^{HP} **Output:** A watermarked high-pass subband $\tilde{C}_{s_w, d_w}^{HP}$

```

1:  $\tilde{C}_{s_w, d_w}^{HP} \leftarrow C_{s_w, d_w}^{HP}$ 
2: for  $j = 0$  to  $\lfloor H/M \rfloor - 1$  do
3:    $E_{s_w, d_w}(j) \leftarrow 0$ 
4:   for  $l = 1$  to  $\mathbb{L}$  do
5:     for  $i = 0$  to  $\lfloor W/N \rfloor - 1$  do
6:        $\tilde{C}_{s_w, d_w}^{HP}(i, j) \leftarrow \lfloor (C_{s_w, d_w}^{HP}(i, j) \times \mathbb{S}) / \Delta^l \rfloor \times \Delta^l$ 
7:        $E_{s_w, d_w}(j) \leftarrow \sum_{i'=0}^{\lfloor W/N \rfloor - 1} |\tilde{C}_{s_w, d_w}^{HP}(i', j) - C_{s_w, d_w}^{HP}(i', j) \times \mathbb{S}|$ 
8:       if  $(E_{s_w, d_w}(j) > \mathbb{E})$  or  $(l == \mathbb{L})$  then
9:         for  $i = 0$  to  $\lfloor W/N \rfloor - 1$  do
10:           $\tilde{C}_{s_w, d_w}^{HP}(i, j) \leftarrow \tilde{C}_{s_w, d_w}^{HP}(i, j) / \mathbb{S}$ 
11:        end for
12:        goto 1: {Branch for quantization of  $(j + 1)$ -th row}
13:      end if
14:    end for
15:  end for
16: end for

```

Algorithm 1 describes the details of the quantization-based watermark embedding. As inspired by [13], [49], the j -th row is quantized by quantizing each coefficient while changing the i index. The quantized coefficient can be obtained from the following equation:

$$\tilde{C}_{s_w, d_w}^{HP}(i, j) = \lfloor (C_{s_w, d_w}^{HP}(i, j) \times \mathbb{S}) / \Delta^l \rfloor \times \Delta^l, \quad (13)$$

where \mathbb{S} and Δ denote scaling factor of coefficients for making finer quantization and size of base quantization step for quantizer, respectively, and l denotes the current quantization level. If the value of l is increased, the quantization step is increased, and thus the embedding strength of the watermark is increased. If the coefficient at the (i, j) coordinate is quantized, a checking process for securing the watermark extraction performance, robustness, and invisibility is followed.

For this, we compared the quantization error and quantization level with experimentally determined thresholds (see line 8 in Algorithm 1). First, we checked the degree of quantization by calculating the quantization error, which is the sum of the difference between the original coefficients and the quantized coefficients [48], [49]. Meanwhile, when the quantization level l is equal to the maximum quantization level, which is denoted as \mathbb{L} , it is evaluated that sufficient watermark embedding has proceeded, and the quantization on the current row of coefficients is then completed. The quantization error of the j -th row of coefficients is denoted as $E(j)_{s_w, d_w}$ and it can be computed by:

$$E_{s_w, d_w}(j) = \sum_{i'=0}^{\lfloor W/N \rfloor - 1} |\tilde{C}_{s_w, d_w}^{HP}(i', j) - C_{s_w, d_w}^{HP}(i', j) \times \mathbb{S}|. \quad (14)$$

The computed $E_{s_w, d_w}(j)$ is compared to the value of the threshold \mathbb{E} , which is determined based on λ_{PM_B} generated by taking the visual characteristics into account as follows:

$$\mathbb{E} = \begin{cases} \omega_1, & \text{if } \lambda_{PM_B} \geq 0.6 \\ \omega_2, & \text{else,} \end{cases} \quad (15)$$

where ω_1 and ω_2 denote the thresholds to control the quantization error and these two parameters are determined considering imperceptibility and robustness. Since B with $\lambda_{PM_B} \geq 0.6$ has inherent properties of content that is less sensitive to watermark embedding, we made the proposed method strongly quantize the coefficients using ω_1 . Given \mathbb{E} , our method can make sufficient statistical difference between the high-pass subbands corresponding to \mathbf{G}_0 and \mathbf{G}_1 and a detailed description of this follows in the next subsection for watermark extraction. If the calculated $E(j)_{s_w, d_w}$ is greater than \mathbb{E} , the quantization for the current row is completed and the result of dividing the temporarily watermarked coefficient value by \mathbb{S} is determined as $\tilde{C}_{s_w, d_w}^{HP}(i, j)$. As illustrated in line 12 of Algorithm 1, the value of j is incremented by 1, followed by quantization for the $(j + 1)$ -th row of coefficients. In the opposite case (i.e. when the condition of line 8 in Algorithm 1 is not satisfied), the value of l is increased to quantize it again for the j -th row of coefficients.

Applying the procedures described in Algorithm 1 to all high-pass subbands corresponding to the group determined by b completes watermark embedding for a given B . The watermarked subblock \tilde{B} is reconstructed by inverse NSCT from the watermarked and original subbands. After watermarking $N \times M$ subblocks, each reconstructed subblock is merged to generate a watermarked Y-channel image. Finally, the watermarked center image, represented by \tilde{I} , can be obtained by converting the YUV representation into an RGB representation.

B. WATERMARK EXTRACTION

Fig. 11 depicts the overall process of watermark extraction; it consists of the following steps.

- 1) Y-channel acquisition and subblock division
- 2) Conduction of three-level NSCT decomposition
- 3) Pairing of high-pass subbands and applying quantization to selected subbands
- 4) Statistical difference computation on paired high-pass subbands
- 5) Watermark bit extraction
- 6) Repeat steps 2-5 for the entire subblock
- 7) Merging the extracted watermark bits
- 8) Unshuffling and multi-bit message extraction

As summarized above, the first two steps are the same as those of the watermark embedding. Since the proposed watermarking method is robust against DIBR attacks, it can extract embedded watermarks from both a watermarked center image and watermarked left or right images. This subsection gives a description on the assumption that a given image denoted \tilde{I} is watermarked center, left and right images. Like the

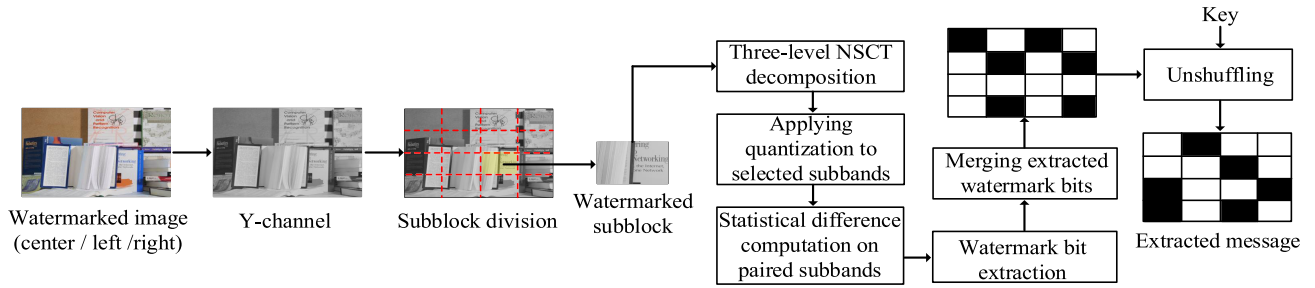


FIGURE 11. Block diagram of the watermark extraction process for the proposed NSCT-based robust and perceptual watermarking for DIBR 3D images.

watermark embedding process, \tilde{I} is converted into Y-channel representation and then divided into $N \times M$ subblocks, where each watermarked subblock is $\lfloor \frac{W}{N} \rfloor \times \lfloor \frac{H}{M} \rfloor$ in size. Let $\tilde{B}_{u,v}$ denote a divided subblock in Y-channel representation where $u \in \{0, 1, \dots, N-1\}$ and $v \in \{0, 1, \dots, M-1\}$. From now on, for the sake of simplicity, we describe the process of extracting the watermark bit from an arbitrary watermarked subblock, represented by \tilde{B} . The three-level NSCT decomposition is applied for \tilde{B} and watermark bit extraction is then conducted on carefully selected high-pass subbands utilized in the watermarking. Here, each high-pass subband generated from \tilde{B} is represented by $\tilde{C}_{s_w, d_w}^{HP}$ where s_w and d_w denote the scale and direction index.

Since quantization is performed on the high-pass subbands of one group among \mathbf{G}_0 and \mathbf{G}_1 , the statistical difference occurs between the two groups after watermarking is complete. We analyze the statistical difference between the two groups by classifying the high-pass subbands that correspond to \mathbf{G}_0 or \mathbf{G}_1 into pairs: $(\tilde{C}_{2,0}^{HP}, \tilde{C}_{2,1}^{HP})$, $(\tilde{C}_{3,0}^{HP}, \tilde{C}_{3,3}^{HP})$, and $(\tilde{C}_{3,1}^{HP}, \tilde{C}_{3,2}^{HP})$, considering that paired subbands have similar properties of scale and direction index. Then, we apply quantization as described in the previous subsection to the coefficients of six selected high-pass subbands, thereby analyzing the differences between the groups to extract the embedded watermark bit. Here, the result of applying Algorithm 1 of quantization to $\tilde{C}_{s_w, d_w}^{HP}$ is represented by $\tilde{C}_{s_w, d_w}^{HP}$. The difference between the coefficients corresponding to the j -th row of $\tilde{C}_{s_w, d_w}^{HP}$ and $\tilde{C}_{s_w, d_w}^{HP}$, can be computed as follows:

$$D_{s_w, d_w}(j) = \sum_{i=0}^{\lfloor \frac{W}{N} \rfloor - 1} |\tilde{C}_{s_w, d_w}^{HP}(i, j) - \tilde{C}_{s_w, d_w}^{HP}(i, j)|. \quad (16)$$

Here, D_{s_w, d_w} and D_{s_w1, d_w1} denote the computed results for subbands corresponding to \mathbf{G}_0 and \mathbf{G}_1 , respectively. We estimate which group is quantized by analyzing the statistical differences based on the calculated D_{s_w, d_w} and D_{s_w1, d_w1} , and then extract an estimated watermark bit \tilde{b} by aggregating the analyzed results using a majority voting system.

Algorithm 2 describes the watermark extraction process in detail. Each pair described above is represented by $\tilde{P}_p = (\tilde{C}_{s_w^0(p), d_w^0(p)}^{HP}, \tilde{C}_{s_w^1(p), d_w^1(p)}^{HP})$ where $p = \{0, 1, 2\}$; it represents an index that refers to each pair. The first and second elements

Algorithm 2 Statistical Difference-Based Watermark Bit Extraction on Given Block

Input: Paired high-pass subbands \tilde{P}_p where $p = \{0, 1, 2\}$
Output: A estimated watermark bit \tilde{b} for given a sub-block

```

1:  $\Theta_0 \leftarrow 0, \Theta_1 \leftarrow 0$ 
2: for  $p = 0$  to  $2$  do
3:    $\tilde{C}_{s_w^0, d_w^0}^{HP} \leftarrow \tilde{C}_{s_w^0(p), d_w^0(p)}^{HP}, \tilde{C}_{s_w^1, d_w^1}^{HP} \leftarrow \tilde{C}_{s_w^1(p), d_w^1(p)}^{HP}$ 
4:   Compute  $\tilde{C}_{s_w^0, d_w^0}^{HP}$  and  $\tilde{C}_{s_w^1, d_w^1}^{HP}$  applying Algorithm 1 to  $\tilde{C}_{s_w^0, d_w^0}^{HP}$  and  $\tilde{C}_{s_w^1, d_w^1}^{HP}$ 
5:   for  $j = 0$  to  $\lfloor H/M \rfloor - 1$  do
6:     Compute  $D_{s_w^0, d_w^0}(j)$  and  $D_{s_w^1, d_w^1}(j)$  using Eq. (16)
7:     if  $(D_{s_w^1, d_w^1}(j) - D_{s_w^0, d_w^0}(j) > \mu)$  then
8:        $\Theta_0 \leftarrow \Theta_0 + 1$ 
9:     else if  $(D_{s_w^0, d_w^0}(j) - D_{s_w^1, d_w^1}(j) > \mu)$  then
10:       $\Theta_1 \leftarrow \Theta_1 + 1$ 
11:    end if
12:  end for
13: end for
14: if  $(\Theta_0 > \Theta_1)$  then
15:    $\tilde{b} \leftarrow 0$ 
16: else
17:    $\tilde{b} \leftarrow 1$ 
18: end if

```

of the parentheses corresponding \tilde{P}_p mean the high-pass subbands that correspond to \mathbf{G}_0 and \mathbf{G}_1 , respectively. The details of scale and direction index for watermark extraction are as follows: $s_w^0 = \{2, 3, 3\}$, $d_w^0 = \{0, 0, 1\}$, $s_w^1 = \{2, 3, 3\}$, and $d_w^1 = \{1, 3, 2\}$. First, $\tilde{C}_{s_w^0, d_w^0}^{HP}$ and $\tilde{C}_{s_w^1, d_w^1}^{HP}$ are computed by applying Algorithm 1 to $\tilde{C}_{s_w^0, d_w^0}^{HP}$ and $\tilde{C}_{s_w^1, d_w^1}^{HP}$, respectively, as specified in lines 3 and 4 of Algorithm 2. Then, $D_{s_w^0, d_w^0}$ and $D_{s_w^1, d_w^1}$ are calculated using Equation (16). In this step, the watermarked and unwatermarked subbands have a low value and a high value for D_{s_w, d_w} , respectively, and we estimate the watermark bit by focusing on this statistical difference in paired high-pass subbands. The conditions of lines 7 and 9 in Algorithm 2 can be used to estimate which group of \mathbf{G}_0 or \mathbf{G}_1 is quantized, where μ represents a threshold for estimating an energy difference due to quantization.

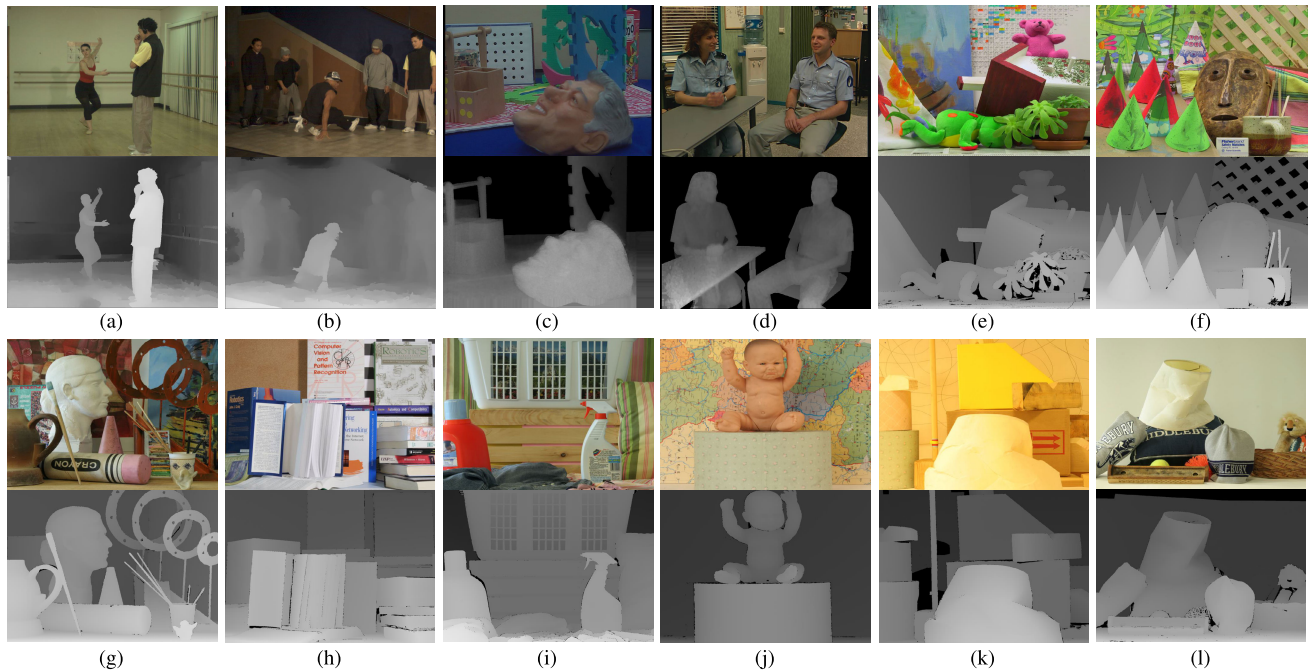


FIGURE 12. Examples of pairs of center and depth images: (a) *Ballet*, (b) *Breakdancers*, (c) *Orbi*, (d) *Interview*, (e) *Teddy*, (f) *Cones*, (g) *Art*, (h) *Books*, (i) *Laundry*, (j) *Baby*, (k) *Lampshade*, and (l) *Midd*.

If it is estimated that the row of coefficients of subbands corresponding to \mathbf{G}_0 is quantized (i.e. if the condition of line 7 is satisfied), Θ_0 is incremented by 1, and in the opposite case, Θ_1 increases by 1. In addition, if the difference between the two calculated statistical differences is greater than 0 and less than μ , it is excluded from counting, because there is not enough evidence to determine which group has been quantized. The processes described so far are repeated for the entire row of coefficients and all pairs. As depicted in lines 14 to 18 Algorithm 2, we employ the majority voting system to extract \tilde{b} . If the value of Θ_0 is greater than the value of Θ_1 , the watermark bit \tilde{b} is estimated to be 0. In the opposite case, \tilde{b} is estimated to be 1. As can be seen in the right part of Fig. 11, watermark bit estimation is applied to all subblocks $\tilde{B}_{u,v}$, where $u \in \{0, 1, \dots, N-1\}$ and $v \in \{0, 1, \dots, M-1\}$, and then the total of $N \times M \tilde{b}$ is estimated. Finally, extracted message \mathbf{M}_e can be obtained after unshuffling.

VI. EXPERIMENTAL RESULTS

We conducted a series of experiments to evaluate the performance of the proposed watermarking method. The experimental data and evaluation criteria are introduced first, and then details of the parameter determination of the proposed method and comparative methods are presented. Finally, we present the experimental results and comparison analysis with existing comparative methods in terms of fidelity and robustness. For fidelity, we performed various metric-based objective fidelity and subjective fidelity tests in which the subject participated. For robustness, after applying signal processing operations, geometric distortions, and DIBR attacks to the watermarked images, we evaluated the degree

to which the extracted message was similar to the original message.

A. EXPERIMENTAL DATA

To demonstrate the effectiveness of the proposed method, a series of experiments were conducted on DIBR datasets consisting of pairs of center and corresponding depth images. Performance evaluation was conducted using the following datasets: Microsoft Research 3D Video Datasets [17], Heinrich-Hertz Institut [4], and Middlebury Stereo Datasets [18]–[20]. Six hundred pairs of color center and grayscale depth images were employed in the experiments with various resolutions from 720×576 to 1800×1500 . We conducted experiments on 200 pairs of images obtained from each of the datasets, respectively. Fig. 12 shows the example of the pairs of center and depth images on the datasets. For Microsoft Research 3D Video Datasets containing '*Ballet*' and '*Breakdancers*' in [17], the resolution of the center and depth images of the dataset was 1024×768 . The dataset consisted of color and depth images generated by shooting eight cameras for two scenes, and we used 200 randomly selected pairs of images for the experiments.

The experimental data, '*Orbi*' and '*Interview*' with 720×576 resolution, can be obtained from Heinrich-Hertz Institut [4], and we randomly extracted 200 pairs from two pieces of sequence data and used them in the experiment. The pairs shown in Fig. 12(e) to Fig. 12(l), '*Teddy*', '*Cones*', '*Art*', '*Books*', '*Laundry*', '*Baby*', '*Lampshade*', and '*Midd*', are examples belonging to Middlebury Stereo Datasets [18]–[20], and we utilized full-sized images with

TABLE 2. Parameter setting of the proposed watermarking method and existing comparative methods for comparative experiments.

Method	Capacity	Parameter and value
IR-based [12]	32 bits	$\alpha = 1, \lambda = 1, B_{size} = 16 \times 16, K_{16 \times 16} = 80, t_x = 5\% \text{ of } W, E_{position} = 10$
DTCWT-based [13]	32 bits	$B_{size} = W/8 \times H/4, errMin = 450, maxBit = 8, \mathbb{W} = 2, TH_1 = 0.0982, TH_2 = -0.0551$
SIFT-based [14]	12 bits	$\alpha = 0.8, K = 4, \theta_S = 45^\circ, \theta_K = 90^\circ, \theta_E = 2^\circ, \theta_E^* = 1.5^\circ, P_h = 32, P_w = 32, M_p = 15, L_w = 320, \rho = 2, E_{position} = 120$
Proposed method	32 bits	$N = 8, M = 4, \alpha = 1, \beta = 0.7, \theta = 0.5, \epsilon_1 = 0.2, \epsilon_2 = 0.8, \eta_1 = 0.6, \eta_2 = 0.4, \mathbb{L} = 8, \mathbb{S} = 2, \Delta = 2, \omega_1 = 400, \omega_2 = 340, \mu = 0.1522$

resolutions ranging from 1240×1110 to 1800×1500 . We only used pairs of views where a depth image existed among multiple views taken for a particular scene. For the 2005 and 2006 datasets [19], [20], there were the versions of combinations of the illumination and exposure parameters for each pair. In the experiments, we used a total of 200 pairs for the Middlebury Stereo Datasets, excluding parameter combinations that had too bright or dark brightness values. As shown in Fig. 12, the experimental images obtained from the datasets [4], [17]–[20] varied in resolution, number of objects, texture, and color.

B. PARAMETER DETERMINATION

In this subsection, details of parameter setting for the DIBR process and comparative experiments are presented. First, the parameters of the DIBR process for generating synthesized left and right images are described. Because the recommended disparity between the synthesized left and right images ranged from 3% to 5% of the width of center image W to offer a comfortable viewing experience to viewers [12]–[14], the baseline distance t_x for adjusting the parallax of two synthesized views was set to 5% of W . The parameters Z_n, Z_f , and f used on the DIBR process were set to 1, $\frac{t_x}{2}$, and 1, respectively. For natural synthesized image generation, an asymmetric Gaussian filter with $\sigma_h = 10$ and $\sigma_v = 30$ was utilized for preprocessing the depth image. The preprocessing was used only for the robustness experiment on smoothed depth image-based view generation, and we utilized the original depth image for other experiments. In addition, linear interpolation was used in the hole-filling process to fill newly exposed holes. Based on these parameters, left and right images were synthesized for the experiments, and the synthesized images were then used for fidelity and robustness tests. In particular, for the robustness test for baseline distance adjustment, we conducted an experiment with varying t_x from 3% to 7% of W .

To show the effectiveness of the proposed method, we designed comparative experiments to analyze the performance of our work versus IR-based [12], DTCWT-based [13], and SIFT-based [14] methods in terms of fidelity and robustness. The proposed method and comparative methods in [12]–[14] were blind watermarking algorithms that could extract watermarks without the preserved side information and original cover work. To conduct experiments under the same

conditions, the parameters of the IR-based, DTCWT-based, and proposed methods were determined to have sufficient watermark capacity (32 bits). The SIFT-based method, which has a limitation in the watermark capacity, was exceptionally set to have a capacity of 12 bits. For all comparative methods, watermarks were embedded into the Y-channel of center images. The details of parameter setting for the proposed and comparative methods are given in Table 2. As depicted in [12], there were two different settings for the IR-based method, depending on the block size B_{size} employed for watermarking. In the experiments, we used a setting of the IR-based method utilizing B_{size} of 16×16 , which guaranteed higher robustness than that of B_{size} of 8×8 . The IR-based method extracted the watermark bit for each block and then generated 32 bits of a message by applying a majority voting system on the basis of predefined intervals.

In the case of DTCWT-based and SIFT-based methods, the key parameters were set based on descriptions in [13], [14]. In particular, the SIFT-based method specializing in imperceptibility and robustness against geometric attacks was parameterized to have a watermark capacity of 12 bits in consideration of the trade-off between capacity and performance. As stated in [14], the SIFT-based method has optimal performance under the parameter conditions listed in Table 2. By comparing the invisibility performance of the proposed method with that of the SIFT-based method, we attempted to verify the effectiveness of the perceptual masking methods exploited in the watermark embedding process. For the proposed method, the values of parameters were empirically determined with consideration of robustness and imperceptibility. In detail, the given Y-channel image was divided into 8×4 subblocks, where each subblock was $\lfloor \frac{W}{8} \rfloor \times \lfloor \frac{H}{4} \rfloor$ in size. The parameters for generating the perceptual masking value λ_{PM_B} were set as $\alpha = 1, \beta = 0.7, \theta = 0.5, \epsilon_1 = 0.2, \epsilon_2 = 0.8, \eta_1 = 0.6$, and $\eta_2 = 0.4$. The parameters of quantization-based watermark embedding and extraction were set as $\mathbb{L} = 8, \mathbb{S} = 2, \Delta = 2, \omega_1 = 400, \omega_2 = 340$, and $\mu = 0.1522$. For the comparative experiments based on the described parameters, we used a desktop computer with an Intel(R) i7-4790K CPU and 16 GB main memory. Each algorithm was implemented in MATLAB R2014a, and the criteria for evaluating the performance in terms of fidelity and robustness are introduced in the following subsection.

C. EVALUATION CRITERIA

We conducted comparative experiments in terms of fidelity and robustness. The fidelity test evaluates the quality of the watermarked image, and we conducted the following two tests: objective fidelity and subjective fidelity tests. For the objective fidelity test, image quality assessment (IQA) metric-based quantitative analysis was performed. Including the peak-signal-to-noise ratio (PSNR) and structural similarity (SSIM) [50], which are most actively used for the evaluation criteria of image quality, we analyzed the quality of watermarked images using the following IQA metrics: feature similarity (FSIM) [51] and multi-scale SSIM (MS-SSIM) [52]. Based on these various IQA metrics, objective analysis between the watermarked and original images was performed.

Next, the subjective fidelity test is a qualitative analysis based on human judgment, and the mean opinion score (MOS), which is popular indicator of perceived media quality is used for subject quality evaluation [53], [54]. In this experiment, a modified version of the double-stimulus continuous quality scale (DSCQS) presentation structure based on [55] was utilized to evaluate the subtle watermark signal centrally. As illustrated in Fig. 13, the original and watermarked images were repeatedly displayed according to the DSCQS presentation structure. Here, O_o and W_o mean original and watermarked images were displayed at the observing time, and O_v and W_v mean original and watermarked images were displayed at the voting time, respectively. At voting time, the subjects assessed the quality of the image on the screen according to the rating scale of the MOS depicted in the right part of Fig. 13. Most often judged on a scale of 1 (Very annoying) to 5 (Imperceptible), the MOS is computed as the average of ratings performed by human subjects for a given stimulus. In summary, the subjects assessed the relative degradation of the watermarked image compared with the original image. On analyzing the subjective fidelity, we focused on the fact that when the MOS value was high, the difference between the original and watermarked images recognized by HVS was small.

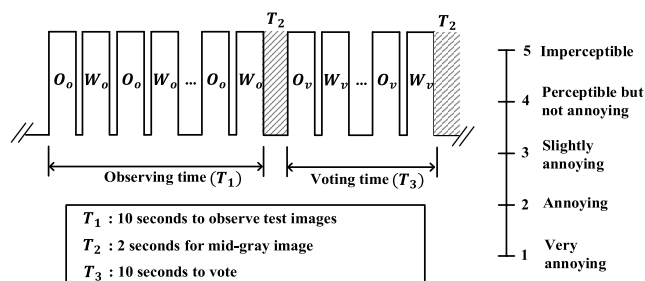


FIGURE 13. The extended version of the DSCQS presentation structure and rating scales used to assess image quality.

Finally, to evaluate the robustness of the watermarking methods, the bit error rate (BER) for computing the amount of difference between original message \mathbf{M}_o and extracted message \mathbf{M}_e was used as an evaluation criterion where

\mathbf{M}_o and \mathbf{M}_e consisted of $N \times M$ bits. Since watermarked images can be exposed to various attacks during transmission and distribution, watermarking methods should be designed to extract \mathbf{M}_e similarly to \mathbf{M}_o even when attacks are applied before watermark extraction. The BER of \mathbf{M}_o and \mathbf{M}_e was equal to $\frac{1}{N \times M} \sum_{i=1}^{N \times M} |b_i^o - b_i^e|$, where b_i^o and b_i^e represent the i -th bit constituting \mathbf{M}_o and \mathbf{M}_e , respectively. Therefore, the more robust the watermarking technique is against an attack, the closer the calculated BER is to 0. To analyze the robustness of the proposed and comparative methods, we conducted the robustness test of each method by applying various attacks, such as DIBR attacks, signal processing operations, and geometric distortions, to the watermarked images and then measuring the BERs.

D. FIDELITY TEST

In this subsection, the objective and subjective quality evaluation results for the watermarked images generated by the proposed and comparable methods are introduced.

1) OBJECTIVE FIDELITY TEST

To evaluate the objective perceptual quality of the watermarking methods, we exploited various IQA metrics, including PSNR, SSIM, FSIM, and MS-SSIM [50]–[52], where PSNR is usually expressed in terms of the logarithmic decibel (dB) scale. We conducted objective fidelity measurements between the watermarked center and original center images based on the IQA metrics. Table 3 shows the experimental results of the objective fidelity test, and the first row of the measures in Table 3 shows the results between the two original center images, which were added for the numerical comparison of each metric. The SIFT-based method showed the best performance because the method embeds watermarks in partial regions based on refined SIFT keypoints. In addition, since the SIFT-based method has a lower watermark capacity of 12 bits, the degree of degradation due to watermark embedding was relatively small compared with that of other methods. The DTCWT-based method, which strongly inserts watermarks into coefficients to achieve high robustness, had the lowest performance for all metrics.

TABLE 3. Results of objective fidelity test of watermarking methods using various IQA metrics.

Method	IQA metric			
	PSNR	SSIM	FSIM	MS-SSIM
Original image	∞	1	1	1
IR-based [12]	43.76	0.9954	0.9971	0.9942
DTCWT-based [13]	41.51	0.9905	0.9918	0.9937
SIFT-based [14]	46.87	0.9972	0.9986	0.9983
Proposed method	45.02	0.9969	0.9986	0.9977

On the other hand, the proposed method aims to improve invisibility through the adjustment of the embedding strength based on the perceptual masking value, the use of selected high-pass subbands, and empirically determined



FIGURE 14. Results of watermarked center image of the proposed and comparative methods. From top to bottom, the results denote the watermarked images of the IR-based, DTCWT-based, SIFT-based, and proposed method, respectively. The second and third columns are the results of enlarging the areas of the yellow and red boxes of the image of each first column, respectively.

optimized parameters. As shown in Table 3, the PSNR, SSIM, FSIM, and MS-SSIM values of the proposed method were 45.02 dB, 0.9969, 0.9986, and 0.9977, respectively. The proposed method achieved performance close to that of the SIFT-based method specialized for objective quality evaluation even with a higher capacity. In addition, our work showed higher quality measures in the IQA metrics-based objective fidelity test than those of the IR-based and DTCWT-based methods, which have the same watermark capacity of 32 bits. In terms of PSNR, the proposed method achieved 1.26 dB and 3.51 dB higher values than the IR-based and DTCWT-based methods, respectively. From this, it can be concluded that our work has high performance in terms of objective quality evaluation.

2) SUBJECTIVE FIDELITY TEST

We performed the subjective fidelity test in addition to metric-based objective evaluation. Fig. 14 shows the results of the watermarked image of the *Books* image using the proposed and comparative methods. From top to bottom of Fig. 14, the results are the watermarked center images of the IR-based, DTCWT-based, SIFT-based, and proposed methods, respectively. The second and third columns were made to enlarge the partial region of the watermarked images of the first column to evaluate the degradation caused by watermark embedding in detail. The areas in the boxes marked in yellow and red were enlarged and listed in the second and third columns, respectively. Since the subjective fidelity test is performed with human eyes, it is possible to evaluate distortions

caused by watermarking that cannot be measured by objective metrics.

In the case of the IR-based method, the block artifacts due to the watermark embedding of the 16×16 DCT coefficients can be seen in the overall part of the watermarked image. For the DTCWT-based method, there was distortion, such as smearing around the letters of the enlarged image of the area for the yellow-marked box. The SIFT-based method achieved the best performance in the objective fidelity test, but when we looked closely at the enlarged image in the third column, there were noticeable block artifacts. Since the watermark embedding of this technique was performed only on the selected block around the keypoint rather than the whole area, the method showed high performance in the objective metric-based evaluation. However, depending on the inherent characteristics of the image content, block artifacts can be highlighted, as in the example of Fig. 14. On the other hand, there were no noticeable distortions or artifacts in the watermarked results of the proposed method compared with the results of the comparative methods (see the bottom row of Fig. 14).

For a more detailed analysis, we conducted the subjective fidelity test based on the computed MOS values. In this experiment, ten subjects participated to evaluate the fidelity of the watermarked images and a Samsung S27E500C Curved LED 27-inch Monitor was employed as the display device. The twelve center images depicted in Fig. 12 were used as the original image, and the four watermarking algorithms were applied to the original image to generate watermarked images. As depicted in Fig. 13, the DSCQS presentation structure was constructed by placing the original and watermarked images at the intersection, and each subject experimented by evaluating the quality of each displayed image as a score. The results of the subjective fidelity test are listed in Table 4. In the MOS-based subjective fidelity test, which evaluates the quality of watermarked images with a score of 1 to 5, the proposed and IR-based methods achieved the high MOS values with above 4.3. The SIFT-based method, which showed the best performance in objective fidelity test, achieved a MOS value of 4.1416, and the DTCWT-based method showed the worst performance with a MOS value of 3.1. From the objective and subjective fidelity tests, we demonstrated that the proposed method has more higher imperceptibility performance than comparative methods.

TABLE 4. Average MOS values of the watermarking methods for test images displayed at the voting time in the subjective fidelity test.

Method	MOS
IR-based [12]	4.3167
DTCWT-based [13]	3.1
SIFT-based [14]	4.1416
Proposed	4.3333

E. ROBUSTNESS TEST

In this subsection, the results of the robustness test for various attacks, including desynchronization attacks from the DIBR process, signal processing operations, and geometric distortions, are introduced.

1) ROBUSTNESS TO DIBR ATTACKS

For the proposed and comparative methods, watermarks were embedded into a center image, and then the left and right images were synthesized from the DIBR process. To deal with the illegal redistribution illustrated in Fig. 1, watermarking methods for DIBR 3D images should be designed to extract embedded watermarks with a low BER value from center, left, and right images. In the robustness test, we first measured the watermark extraction performance in the watermarked center images. Table 5 shows the average BER values of the watermarking methods for watermarked center images in each dataset, where MSR, HHI, and MSD indicate the abbreviation of Microsoft Research 3D Video Datasets [17], Heinrich-Hertz Institut [4], and Middlebury Stereo Datasets [18]–[20], respectively. As listed in Table 5, all methods showed a low BER below 0.03 for each dataset, and the proposed method achieved BER values of 0, 0, and 0.0127 for MSR, HHI, and MSD, respectively. For the center view, the DTCWT-based method showed the worst performance with an average BER value of 0.0129.

TABLE 5. Average BER values of watermarking methods for watermarked center images in each dataset.

Method	Dataset		
	MSR	HHI	MSD
IR-based [12]	0	0	0.0006
DTCWT-based [13]	0.0118	0.0004	0.0265
SIFT-based [14]	0.0012	0.0138	0.0073
Proposed	0	0	0.0127

Next, we evaluated the watermark extraction performance for synthesized left and right images generated from the 3D image warping of the DIBR process with predefined parameters. As listed in Table 6, the IR-based method showed the lowest BER for the center view but worse performance for the left and right views. The average BER values for the left and right images of the proposed method were 0.0052 and 0.0056, respectively. For 3D image warping from gen-

TABLE 6. Average BER values of watermarking methods for watermarked center, left, and right images.

Method	View		
	Center	Left	Right
IR-based [12]	0.0002	0.0176	0.0192
DTCWT-based [13]	0.0129	0.0217	0.0201
SIFT-based [14]	0.0074	0.0108	0.0103
Proposed	0.0042	0.0052	0.0056

erating left or right views, the proposed work achieved the best performance. Compared with the comparative methods where performance degradation occurred for view synthesis, the proposed method showed little difference in the performance of watermark extraction between the center and synthesized views, and it can be concluded that the proposed method was robust to desynchronization occurring on 3D image warping with predefined t_x . This strength derives from the fact that the proposed method utilizes quantization-based watermark embedding on row coefficients and selected NSCT high-pass subbands from the robustness analysis in Section IV.

In addition, we conducted a robustness test for the following two types of DIBR attacks: baseline distance adjustment and depth image preprocessing. Baseline distance adjustment is used to control the depth condition of synthesized views, and it can be regarded as a desynchronization attack in that it affects the distance of pixels moving in the horizontal direction for view synthesis. In this experiment, t_x was set to range from 3% to 7% of the width of center image W where the baseline distance ratio Φ_{t_x} was defined as $\frac{t_x}{W} \times 100$. As listed in Table 7, the proposed method achieved the best performance (BER values below 0.006) for various Φ_{t_x} ranging from 3 to 7. Although the DTCWT-based and SIFT-based methods have higher BER values than that of the proposed work, two methods achieved acceptable performance because of the watermarking algorithm with consideration of baseline distance adjustment. In contrast, the IR-based method can extract the watermark only in the virtual view generated based on the predefined condition in [12]. Therefore, for the IR-based method, the BER value increases except when the predefined condition (i.e. $\Phi_{t_x} = 5$) is met.

TABLE 7. Average BER values of watermarking methods for various baseline distance ratios.

Method	Φ_{t_x}				
	3	4	5	6	7
IR-based [12]	0.2117	0.1093	0.0176	0.1125	0.2018
DTCWT-based [13]	0.0161	0.0202	0.0217	0.0225	0.0234
SIFT-based [14]	0.0124	0.0105	0.0108	0.0101	0.0111
Proposed	0.0051	0.0052	0.0052	0.0053	0.0051

The preprocessing of the depth image is used for reducing sharp depth discontinuities in the depth image, which also affects the synchronization of watermarks in the process of smoothing the depth image. To test the robustness of the watermarking methods, we preprocessed the depth image using an asymmetric filter and then generated synthesized images under the same conditions as those of the previous experiment on baseline distance adjustment. As listed in Table 8, the proposed approach achieved lower BER than comparable methods. When comparing the results of Tables 7 and 8, it can be seen that the preprocessing of the depth image affected watermark extraction. In some cases, the change in the depth condition was adversely affected,

TABLE 8. Average BER values of watermarking methods with preprocessing of depth map for various baseline distance ratios.

Method	Φ_{t_x}				
	3	4	5	6	7
IR-based [12]	0.2165	0.1197	0.0248	0.1216	0.2081
DTCWT-based [13]	0.0165	0.0188	0.0203	0.0226	0.0238
SIFT-based [14]	0.0115	0.0092	0.0098	0.0117	0.0116
Proposed	0.0049	0.0053	0.0051	0.0052	0.0050

while in other cases, the BER values slightly decreased due to the generation of natural views. For a combination attack of preprocessing of depth image and baseline distance adjustment, our work showed lower BER values than other comparative methods. From the robustness test in this subsection, we demonstrated that the proposed method has more stable and higher watermark extraction performance against three types of DIBR attacks than comparative methods.

2) ROBUSTNESS TO SIGNAL PROCESSING OPERATIONS AND GEOMETRIC DISTORTIONS

The watermarked contents undergo various attacks in the distribution process, and these attacks adversely affect watermark extraction. In this subsection, we present the robustness test results of watermarking methods for signal processing operations and geometric distortions, such as JPEG compression, white Gaussian noise addition, salt and pepper noise addition, cropping, translation, and scaling. To demonstrate the effectiveness of the proposed method, we attempted to extract watermarks from synthesized left images after applying various attacks. The results of the robustness test for signal processing operations and geometric attacks are depicted in Figs. 15 and 16, respectively. Each attack was implemented through functions in MATLAB, and translation indicates a type of geometric attack in which pixels in an image move due to a translation factor and zero padding is applied to the area before movement.

We first conducted the robustness test of watermarking methods against signal processing operations. When the strongest parameters for JPEG compression, white Gaussian noise addition, and salt and pepper noise addition were applied, the average PSNR values were 35.71 dB, 30.79 dB, and 28.92 dB, respectively, indicating severe degradation of the watermarked image. For JPEG compression, four watermarking methods had low BER values below 0.055 when the JPEG quality was within the range of 80 to 100. In the case of JPEG quality below 80, the BER value of the IR-based method rapidly increased. Additionally, the proposed method had robustness close to the best performance results of the SIFT-based method. For white Gaussian noise addition, the proposed and IR-based methods outperformed the DTCWT-based and SIFT-based methods. When the variance of Gaussian noise was 8.0×10^{-4} , the BER values of the proposed and IR-based methods were 0.0524 and 0.0637, respectively. On the other hand, for salt and pepper noise addition, the proposed work outperformed the

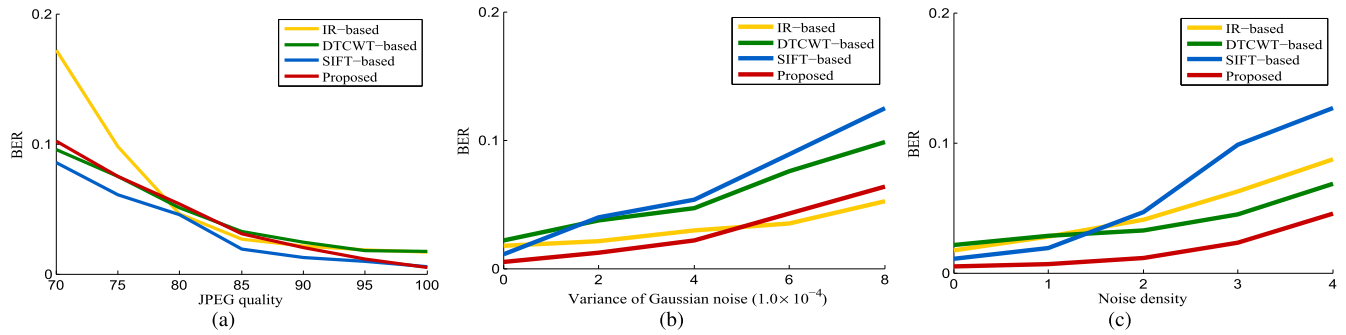


FIGURE 15. Average BER values of watermarking methods for various types of signal processing operations: (a) JPEG compression, (b) white Gaussian noise addition, and (c) salt and pepper noise addition.

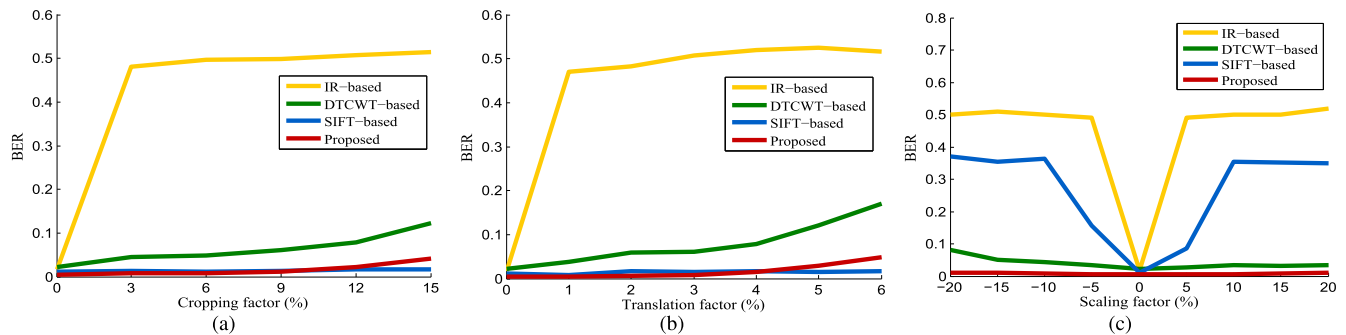


FIGURE 16. Average BER values of watermarking methods for various types of geometric distortions: (a) cropping, (b) translation, and (c) scaling.

IR-based method for noise density in the range of 1 to 4. When the noise density was 4, the BER value of the proposed method was 0.0453. For three types of signal processing operations, the proposed method showed sufficient robustness against the level of attacks that can be applied in real environments.

For geometric distortions, the proposed method had stable and low BER values for three types of geometric attacks, as shown in Fig. 16. For cropping and translation, the BER values of the proposed method slightly increased as the factor value increased, but the BERs were very small values of less than 0.05. Furthermore, the proposed method outperformed all comparative methods for scaling attacks ranging from -20% to 20% . While the proposed method maintained a low BER for various scaling factors, comparative methods showed increases in BER as the scaling factor changed from 0%. The IR-based method achieved the worst performance against geometric attacks, and the DTCWT-based method had a higher BER than the proposed method but showed overall stable and acceptable performance. While the SIFT-based method achieved the best performance against cropping and translation, the method was not robust against scaling that adversely affected the extraction of SIFT keypoints. From the test results in Fig. 16, we confirmed that the proposed method has more outstanding watermark extraction performance against geometric distortions than comparative methods.

VII. CONCLUSION

In this paper, we propose an NSCT-based robust and perceptual watermarking framework for DIBR 3D images. We believe that the NSCT domain is advantageous for securing invisibility and robustness compared to other domains. Compared to the DTCWT domain, the NSCT domain without subsampling in the decomposition process has a little degradation in image quality due to watermark insertion. In addition, because the NSCT domain has property of shift invariance, the NSCT domain was considered suitable for DIBR 3D image watermarking, which was verified through the results of the robustness analysis and various robustness experiments in previous sections. The proposed watermarking framework is the first attempt to utilize NSCT for designing DIBR 3D image watermarking, and we expect that the proposed study will lead to various follow-up studies on NSCT domain-based DIBR 3D image watermarking. In the field of watermarking, it is important not only to select a domain for watermarking, but also to design a framework well by combining each module of watermark insertion and extraction with consideration of robustness and imperceptibility.

For robustness, we have selected directional high-pass subbands that are robust against DIBR attacks, and the proposed watermarking framework inserted and extracted watermarks in rows coefficients of selected subbands to be robust against horizontal desynchronization of DIBR attacks. For imperceptibility, watermark embedding was performed on

high-pass subbands rather than a low-frequency subband, and the strength of watermark embedding was adjusted by considering the texture and brightness characteristics using perceptual masking. To achieve robustness and invisibility at the same time, we carefully selected the parameters of the watermark insertion and extraction modules and integrated them into one framework. Through the comparative experiments, superiority of the proposed watermarking framework was experimentally proved in terms of robustness and imperceptibility. As we intended, our method showed stable and high watermark extraction performance against DIBR attacks, signal processing operations, and geometric distortions, while maintaining high imperceptibility. Additionally, the proposed method achieved higher better performance than the comparative methods in the objective and subjective fidelity tests. In future work, we will extend the proposed watermarking framework for DIBR 3D images to be applicable to video with MVD format.

REFERENCES

- [1] M. Asikuzzaman, M. J. Alam, A. J. Lambert, and M. R. Pickering, "Robust DT-CWT-based DIBR 3D video watermarking using chrominance embedding," *IEEE Trans. Multimedia*, vol. 18, no. 9, pp. 1733–1748, Sep. 2016.
- [2] M. Asikuzzaman and M. R. Pickering, "An overview of digital video watermarking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 9, pp. 2131–2153, Sep. 2018.
- [3] P.-J. Lee, "Nongeometric distortion smoothing approach for depth map preprocessing," *IEEE Trans. Multimedia*, vol. 13, no. 2, pp. 246–254, Apr. 2011.
- [4] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," *Proc. SPIE*, vol. 5291, May 2004, pp. 93–104.
- [5] C. Fehn and R. S. Pastoor, "Interactive 3-DTV-Concepts and key technologies," *Proc. IEEE*, vol. 94, no. 3, pp. 524–538, Mar. 2006.
- [6] L. Zhang and W. J. Tam, "Stereoscopic image generation based on depth images for 3D TV," *IEEE Trans. Broadcast.*, vol. 51, no. 2, pp. 191–199, Jun. 2005.
- [7] N. S. Holliman, N. A. Dodgson, G. E. Favalora, and L. Pockett, "Three-dimensional displays: A review and applications analysis," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 362–371, Jun. 2011.
- [8] J.-H. Cho, W. Song, H. Choi, and T. Kim, "Hole filling method for depth image based rendering based on boundary decision," *IEEE Signal Process. Lett.*, vol. 24, no. 3, pp. 329–333, Mar. 2017.
- [9] M. M. Hannuksela, M. Gabbouj, D. Rusanovskyy, W. Su, L. Chen, R. Li, P. Aflaki, D. Lan, M. Joachimiak, and H. Li, "Multiview-Video-Plus-Depth coding based on the advanced video coding standard," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3449–3458, Sep. 2013.
- [10] D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2366–2374.
- [11] C. Yao, T. Tillo, Y. Zhao, J. Xiao, H. Bai, and C. Lin, "Depth map driven hole filling algorithm exploiting temporal correlation information," *IEEE Trans. Broadcast.*, vol. 60, no. 2, pp. 394–404, Jun. 2014.
- [12] Y.-H. Lin and J.-L. Wu, "A digital blind watermarking for depth-image-based rendering 3D images," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 602–611, Jun. 2011.
- [13] H.-D. Kim, J.-W. Lee, T.-W. Oh, and H.-K. Lee, "Robust DT-CWT watermarking for DIBR 3D images," *IEEE Trans. Broadcast.*, vol. 58, no. 4, pp. 533–543, Dec. 2012.
- [14] S.-H. Nam, W.-H. Kim, S.-M. Mun, J.-U. Hou, S. Choi, and H.-K. Lee, "A SIFT features based blind watermarking for DIBR 3D images," *Multimedia Tools Appl.*, vol. 77, no. 7, pp. 7811–7850, Apr. 2018, doi: 10.1007/s11042-017-4678-x.
- [15] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker, *Digital Watermarking and Steganography*. San Mateo, CA, USA: Morgan Kaufmann, 2007.
- [16] G. C. Langelaar, I. Setyawan, and R. L. Legendijk, "Watermarking digital image and video data. A state-of-the-art overview," *IEEE Signal Process. Mag.*, vol. 17, no. 5, pp. 20–46, Sep. 2000.
- [17] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 600–608, Aug. 2004.
- [18] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
- [19] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
- [20] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jun. 2003, p. 1.
- [21] E. Halici and A. A. Alatan, "Watermarking for depth-image-based rendering," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 4217–4220.
- [22] M.-J. Lee, J.-W. Lee, and H.-K. Lee, "Perceptual watermarking for 3D stereoscopic video using depth information," in *Proc. 7th Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.*, Oct. 2011, pp. 81–84.
- [23] S. Wang, C. Cui, and X. Niu, "Watermarking for DIBR 3D images based on SIFT feature points," *Measurement*, vol. 48, pp. 54–62, Feb. 2014.
- [24] H. Miao, Y.-H. Lin, and J.-L. Wu, "Image descriptor based digital semi-blind watermarking for DIBR 3D images," in *Proc. Int. Workshop Digit. Watermarking*. Cham, Switzerland: Springer, 2014, pp. 90–104.
- [25] C. Cui, S. Wang, and X. Niu, "A novel watermarking for DIBR 3D images with geometric rectification based on feature points," *Multimedia Tools Appl.*, vol. 76, no. 1, pp. 649–677, Jan. 2017.
- [26] J. W. Lee, H.-D. Kim, H.-Y. Choi, S. Choi, and H.-K. Lee, "Stereoscopic watermarking by horizontal noise mean shifting," *Media Watermarking, Secur., Forensics*, vol. 8303, Feb. 2012, Art. no. 830307.
- [27] S. Rana and A. Sur, "Depth-based view-invariant blind 3D image watermarking," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 12, no. 4, pp. 1–23, Aug. 2016.
- [28] C. Cui and X.-M. Niu, "A robust DIBR 3D image watermarking algorithm based on histogram shape," *Measurement*, vol. 92, pp. 130–143, Oct. 2016.
- [29] W.-H. Kim, J.-U. Hou, H.-U. Jang, and H.-K. Lee, "Robust template-based watermarking for DIBR 3D images," *Appl. Sci.*, vol. 8, no. 6, p. 911, 2018. [Online]. Available: <http://www.mdpi.com/2076-3417/8/6/911>
- [30] L. Zhang, C. Vazquez, and S. Knorr, "3D-TV content creation: Automatic 2D-to-3D video conversion," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 372–383, Jun. 2011.
- [31] P. Ndjiki-Nya, M. Koppel, D. Doshkov, H. Lakshman, P. Merkle, K. Müller, and T. Wiegand, "Depth image-based rendering with advanced texture synthesis for 3-D video," *IEEE Trans. Multimedia*, vol. 13, no. 3, pp. 453–465, Jun. 2011.
- [32] W. Tam, G. Alain, L. Zhang, T. Martin, R. Renaud, and D. Wang, "Smoothing depth maps for improved stereoscopic image quality," *Proc. SPIE*, vol. 5599, Oct. 2004, pp. 162–172.
- [33] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.
- [34] A. L. Da Cunha, J. Zhou, and M. N. Do, "The nonsubsampling contourlet transform: Theory, design, and applications," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3089–3101, Oct. 2006.
- [35] G. Bhatnagar, Q. M. J. Wu, and Z. Liu, "Directive contrast based multimodal medical image fusion in NSCT domain," *IEEE Trans. Multimedia*, vol. 15, no. 5, pp. 1014–1024, Aug. 2013.
- [36] S. Singh, V. S. Rathore, and R. Singh, "Hybrid NSCT domain multiple watermarking for medical images," *Multimedia Tools Appl.*, vol. 76, no. 3, pp. 3557–3575, Feb. 2017.
- [37] S. Singh, R. Singh, A. K. Singh, and T. J. Siddiqui, "SVD-DCT based medical image watermarking in NSCT domain," in *Quantum Computing: An Environment for Intelligent Large Scale Real Application*. Cham, Switzerland: Springer, 2018, pp. 467–488.
- [38] K.-L. Hua, B.-R. Dai, K. Srinivasan, Y.-H. Hsu, and V. Sharma, "A hybrid NSCT domain image watermarking scheme," *EURASIP J. Image Video Process.*, vol. 2017, no. 1, p. 10, Dec. 2017.
- [39] C. Kumar, A. Singh, P. Kumar, R. Singh, and S. Singh, "SPIHT-based multiple image watermarking in NSCT domain," *Concurrency Comput. Pract. Exper.*, vol. 32, no. 1, p. e4912, Jan. 2020.
- [40] C. Tian, R.-H. Wen, W.-P. Zou, and L.-H. Gong, "Robust and blind watermarking algorithm based on DCT and SVD in the contourlet domain," *Multimedia Tools Appl.*, vol. 79, nos. 11–12, pp. 7515–7541, Mar. 2020.

[41] X. Wu, J. Li, R. Tu, J. Cheng, U. A. Bhatti, and J. Ma, "Contourlet-DCT based multiple robust watermarks for medical images," *Multimedia Tools Appl.*, vol. 78, no. 7, pp. 8463–8480, Apr. 2019.

[42] L. Tang, F. Zhao, and Z.-G. Zhao, "The nonsampled contourlet transform for image fusion," in *Proc. Int. Conf. Wavelet Anal. Pattern Recognit.*, vol. 1, Nov. 2007, pp. 305–310.

[43] Z. Li, S. Q. Chen, and X. Y. Cheng, "Dual video watermarking algorithm based on SIFT and HVS in the contourlet domain," *IEEE Access*, vol. 7, pp. 84020–84032, 2019.

[44] M. Barni, F. Bartolini, and A. Piva, "Improved wavelet-based watermarking through pixel-wise masking," *IEEE Trans. Image Process.*, vol. 10, no. 5, pp. 783–791, May 2001.

[45] S. Voloshynovskiy, A. Herrigel, N. Baumgaertner, and T. Pun, "A stochastic approach to content adaptive digital image watermarking," in *Proc. Int. Workshop Inf. Hiding*, Berlin, Germany: Springer, 1999, pp. 211–236.

[46] N. Mettripun, T. Amornraksa, and E. J. Delp, "Robust image watermarking based on luminance modification," *J. Electron. Imag.*, vol. 22, no. 3, Aug. 2013, Art. no. 033009.

[47] H. Kim, "Robust image watermarking using local invariant features," *Opt. Eng.*, vol. 45, no. 3, Mar. 2006, Art. no. 037002.

[48] D. Kundur and D. Hatzinakos, "Digital watermarking using multiresolution wavelet decomposition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, vol. 5, May 1998, pp. 2969–2972.

[49] S.-H. Wang and Y.-P. Lin, "Wavelet tree quantization for copyright protection watermarking," *IEEE Trans. Image Process.*, vol. 13, no. 2, pp. 154–165, Feb. 2004.

[50] 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.

[51] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[52] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 2, 2003, pp. 1398–1402.

[53] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2366–2369.

[54] S.-W. Jung, J.-Y. Jeong, and S.-J. Ko, "Sharpness enhancement of stereo images using binocular just-noticeable difference," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1191–1199, Mar. 2012.

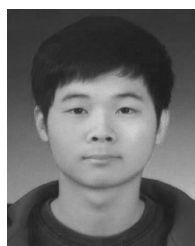
[55] BT, ITU-R. Recommendation, "500-10. Methodology for the subjective assessment of the quality of television pictures," ITU, Geneva, Switzerland, Tech. Rep., 2000.



WONHYUK AHN received the B.S. degree in software and computer engineering from Aju University, South Korea, in 2016. He is currently pursuing the Ph.D. degree with the School of Computing, Korea Advanced Institute of Science and Technology (KAIST). His research interests include multimedia forensics, computer vision, and machine learning.



DONGKYU KIM received the B.S. degree in electronics and communications engineering from Hanyang University, Seoul, South Korea, in 2013, and the M.S. degree in electrical engineering and the Ph.D. degree in computer science from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2015 and 2019, respectively. Since 2019, he has been a Staff Engineer with the Samsung Research, Samsung Electronics Company, Ltd., Seoul, South Korea. His research interests include various aspects of digital image forensics, machine learning, and deep learning.



IN-JAE YU received the B.S. degree from the Department of Mathematical Sciences, Korea Advanced Institute of Science and Technology (KAIST), South Korea, in 2015, and the M.S. degree from the School of Computing from KAIST, in 2017. He is currently pursuing the Ph.D. degree with the Multimedia Computing Laboratory, School of Computing, KAIST. His research interests include digital multimedia forensics and deep learning.



WOOK-HYUNG KIM received the B.S. degree in electrical engineering from Hanyang University, Seoul, South Korea, in 2012, and the M.S. degree in electrical engineering and the Ph.D. degree in computer science from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2014 and 2018, respectively. Since 2018, he has been a Staff Engineer with Samsung Electronics Company, Ltd., Suwon, South Korea. His research interests include various aspects of digital image watermarking, low-level vision, and deep learning.



HEUNG-KYU LEE received the B.S. degree in electronics engineering from Seoul National University, Seoul, South Korea, in 1978, and the M.S. and Ph.D. degrees in computer science from the Korea Advanced Institute of Science and Technology (KAIST), South Korea, in 1981 and 1984, respectively. Since 1986, he has been a Professor with the School of Computing, KAIST. He has authored/coauthored over 200 international journal and conference papers. His major interests are digital watermarking, digital fingerprinting, and digital rights management. He has been a reviewer of many international journals, including the *Journal of Electronic Imaging*, *Real-Time Imaging*, and the *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*.



SEUNG-HUN NAM received the B.S. degree in information communication engineering from Dongguk University, Seoul, South Korea, in 2013, and the M.S. degree from the School of Computing, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2015. He is currently pursuing the Ph.D. degree with the Multimedia Computing Laboratory, School of Computing, KAIST. His research interests include various aspects of digital watermarking, multimedia forensics, and deep learning.



SEUNG-MIN MUN received the B.S. degree from the Department of Mathematical Sciences, Korea Advanced Institute of Science and Technology (KAIST), South Korea, in 2014, and the M.S. and Ph.D. degrees from the School of Computing from KAIST, in 2016 and 2019, respectively. He is currently working with Samsung Electronics Company, Ltd., Suwon, South Korea.