Stereo Vision-Based High Dynamic Range Imaging Using Differently-Exposed Image Pair

Won-Jae Park [1], Seo-Won Ji [1], Seok-Jae Kang [2], Seung-Won Jung [3] and Sung-Jea Ko [1,*]

[1]   School of Electrical Engineering, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 136-701, Korea; wjpark@dali.korea.ac.kr (W.-J.P.); swji@dali.korea.ac.kr (S.-W.J.)
[2]   Samsung Electronics Co. Ltd., 1, Samsungjeonja-ro, Hwaseong-si 445-330, Gyeonggi-do, Korea; seokjae.kang@samsung.com
[3]   Department of Multimedia Engineering, Dongguk University, Pildong-ro 1gil 30, Jung-gu, Seoul 100-715; Korea; swjung83@dongguk.edu
*   Correspondence: sjko@korea.ac.kr; Tel.: +82-2-3290-3228

Abstract: In this paper, a high dynamic range (HDR) imaging method based on the stereo vision system is presented. The proposed method uses differently exposed low dynamic range (LDR) images captured from a stereo camera. The stereo LDR images are first converted to initial stereo HDR images using the inverse camera response function estimated from the LDR images. However, due to the limited dynamic range of the stereo LDR camera, the radiance values in under/over-exposed regions of the initial main-view (MV) HDR image can be lost. To restore these radiance values, the proposed stereo matching and hole-filling algorithms are applied to the stereo HDR images. Specifically, the auxiliary-view (AV) HDR image is warped by using the estimated disparity between initial the stereo HDR images and then effective hole-filling is applied to the warped AV HDR image. To reconstruct the final MV HDR, the warped and hole-filled AV HDR image is fused with the initial MV HDR image using the weight map. The experimental results demonstrate objectively and subjectively that the proposed stereo HDR imaging method provides better performance compared to the conventional method.

---

1. Introduction

Most commercial charge coupled device (CCD) or complementary metal-oxide semiconductor (CMOS) sensors deliver a limited dynamic range (DR) which is usually several orders of magnitude lower than that of a real scene. To overcome such limitation of image sensors, many researchers have developed various DR extension methods, which are also called high dynamic range (HDR) imaging. While some approaches enhance the DR by using particular sensors [1–3], other HDR imaging methods use image processing techniques to generate a high-quality HDR image from low dynamic range (LDR) images captured by low-cost cameras. These HDR imaging methods use multiple LDR images of the same scene captured under different exposures and fuse them into the HDR image [4–7]. Thus, the resultant HDR image has a wide DR similar to a real scene. However, since the fusion process assumes that the scene is completely static, the faint appearance of objects, called ghosting artifact, is often observed in the final HDR image especially when the scene contains moving objects. Jacob and coworkers' method [8] attempts to reduce the ghosting artifact by explicitly detecting and removing the moving objects when combining multi-exposed images, but moving object detection itself is challenging especially for multi-exposed images. Other ghost removal algorithms [9,10] use the PatchMatch-based global optimization to obtain the artifact-free HDR image, but the algorithms

generally require high computational cost. Although the PatchMatch-based global optimization can be accelerated by Lu and coworkers' method [11], which uses the super-pixel as a basic unit for the PatchMatch, super-pixel segmentation itself is challenging for the images captured under the limited dynamic range. The ghosting artifact can also be removed by using special sensors that support spatially varying pixel exposures [12]. However, such sensors not only increase the cost of the cameras but also reduce the resolution of the resultant HDR image.

In an attempt to deal with the ghosting artifact using general sensors, several researchers have focused on reconstructing an HDR image from a stereo image pair captured with different exposures, called stereo HDR imaging [13–16]. According to the literature survey [17], these methods [13–16] correspond to the single source method because a single input image is used from each camera. One method [13] uses the left HDR image and the right LDR image to generate the right HDR image, and the others [14–16] use a stereoscopic LDR image pair as an input. Since the input LDR image pair is captured together with slightly different exposures, the ghosting artifact can be largely alleviated in the resultant HDR image. However, the input LDR images of the stereo HDR imaging system have different viewpoints with each other. Thus, in stereo HDR imaging, two challenging issues have to be addressed: First, high-performance disparity estimation is required to align the AV with the MV. Second, an effective hole-filling method has to be employed to restore the radiance values of the hole regions in the HDR image.

The proposed HDR imaging method adopts a basic framework of the conventional stereo HDR imaging methods [14–16] which consist of image rectification, radiance space conversion, disparity estimation, image warping, and image fusion. In the proposed method, we reconstruct the final HDR image with considering the aforementioned two challenging issues. First, the human visual system (HVS)-based cost computation and segmentation-based cost aggregation method are proposed to obtain a more precise disparity map. Second, the effective hole-filling method is proposed to compensate for the incomplete region in the warped AV HDR image.

The remainder of this paper is organized as follows. The proposed stereo HDR imaging is described in Section 2. In Section 3, experimental results are presented and discussed. Finally, this paper is concluded in Section 4.

2. Proposed Stereo HDR Imaging Method

2.1. Overall Framework

An overall framework for the proposed stereo HDR imaging method is shown in Figure 1. The framework of the proposed method is similar to the flow of conventional stereo HDR imaging methods [14–16]. First, a stereo image pair is transformed into a common image plane using the well-known rectification method [18] prior to the stereo matching process. A stereo LDR image pair of the MV image I_m and the AV image I_a is assumed to be horizontally aligned by rectification. Next, both I_m and I_a are used to obtain the inverse camera response function (ICRF) [19]. After obtaining the initial HDR images by the radiance space conversion, the resultant HDR images of the proposed stereo HDR imaging method are obtained by three major sub-processes: (1) disparity estimation; (2) warping and hole-filling; and (3) fusion with the warped and hole-filled AV HDR image. The segmentation results of the initial HDR images are also used as supplementary information in the first two sub-processes, as shown in Figure 1.
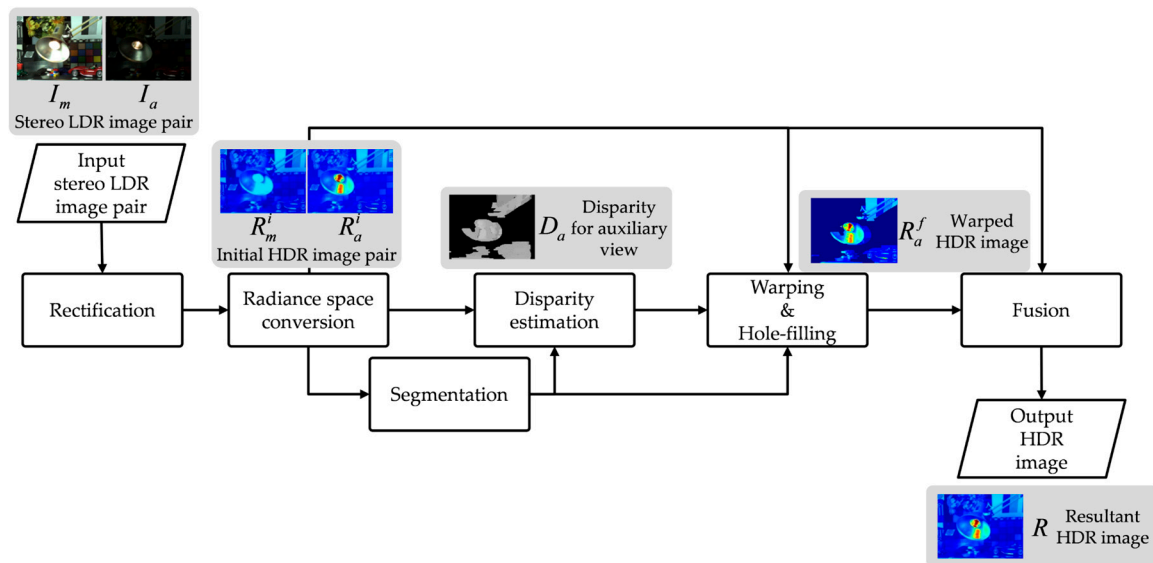
Figure 1. A framework of the proposed stereo high dynamic range (HDR) imaging.

However, since the camera response function (CRF) estimation method [19] is presented for the single-view images with different exposure times, called the bracketed images, the method should be modified for the stereo HDR imaging system. To apply the CRF estimation method [19] to the stereo HDR imaging system, it is mandatory to obtain the pixel correspondences between the stereo images, which are called sample points [14]. The conventional stereo HDR imaging method [14] employs the SIFT descriptor to find the sample points. However, the conventional method [14] does not guarantee the sample points for the full intensity range due to using the SIFT descriptor. To obtain enough sample points for the entire intensity range, the proposed HDR imaging method utilizes the cumulative distribution functions (CDFs) of the stereo images. The proposed method assumes that the pixel values with the same probability at two different CDFs of the stereo images have the similar irradiance value. Thus, given the two CDFs, the sample points are collected by selecting the pairs of pixel values that correspond to the same probability. For example, the intensity value 18 of the left image is matched with the intensity value 113 of the right image, as shown in Figure 2a. All the collected sample points are then used to estimate the ICRF [19]. Figure 2b–d shows the resulting ICRF curves of the conventional method [14], the proposed method, and the reference method [19]. The ICRF curves of the conventional and the proposed method are obtained by using the stereo images with two different exposure times (short and long). The reference ICRF [19] is generated by using the single-view images with the three different exposure times (short, normal, and long). It can be seen that the ICRFs of the proposed ICRF method for the stereo HDR imaging system are more accurate than the conventional ones [14].

Figure 3 shows a radiance space conversion process. The rectified LDR images are first converted to the initial HDR images, R_m^i and R_a^i, by using the proposed ICRF method. These initial HDR images are then used for the proposed stereo HDR imaging system.
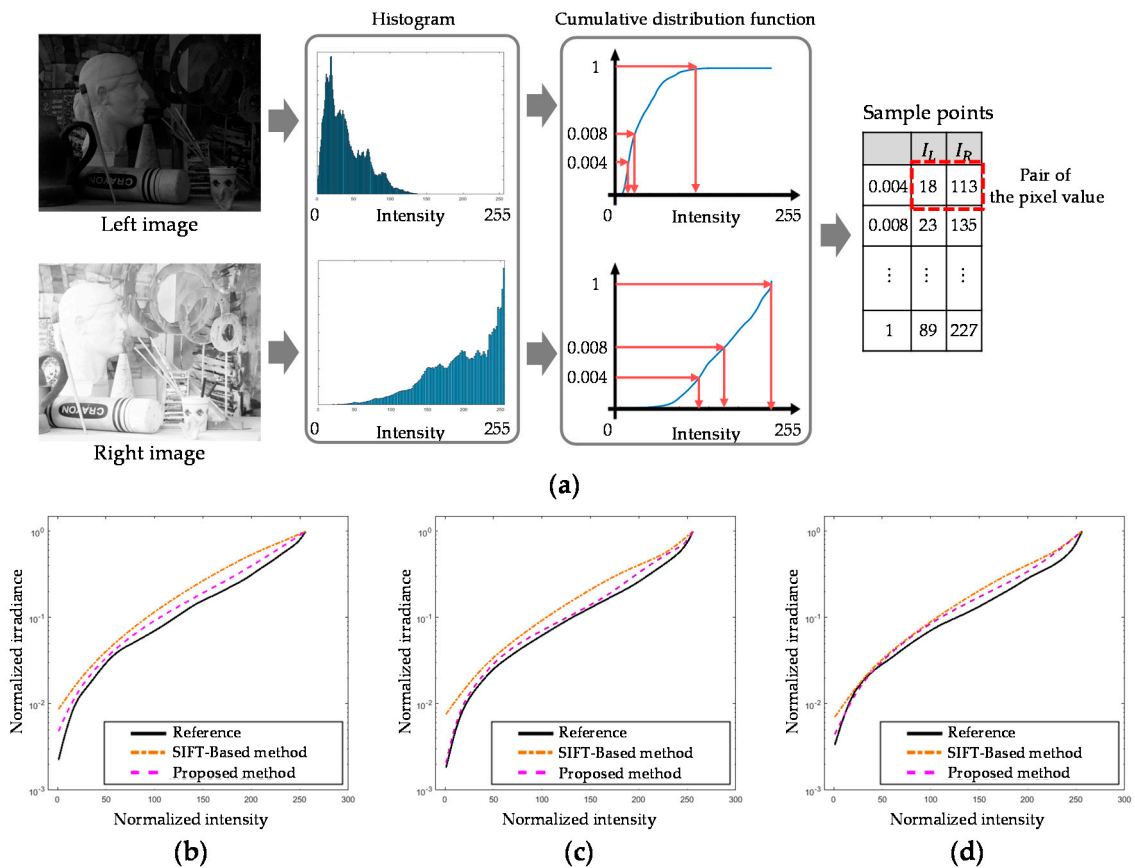
Figure 2. (**a**) An example of the proposed method; and the obtained inverse camera response functions (ICRFs) of: (**b**) the red channel; (**c**) the green channel; and (**d**) the blue channel.
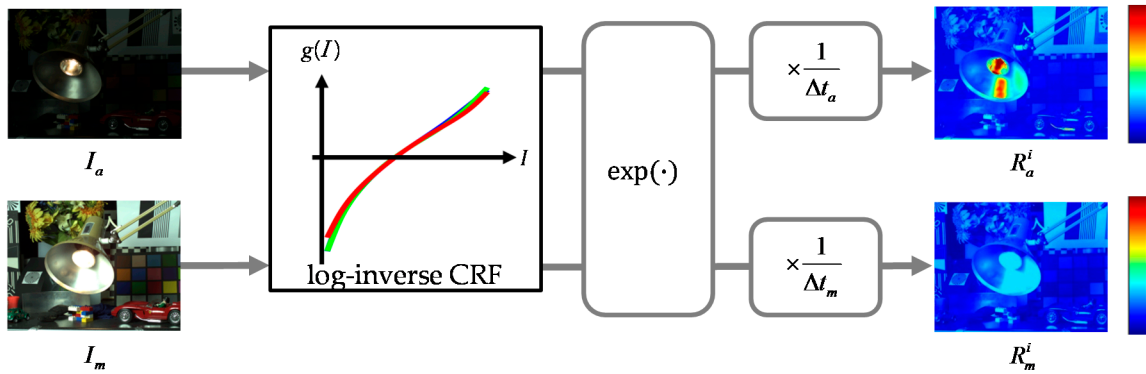

Figure 3. Radiance space conversion.

Similar to the conventional algorithms [14–16], the left-view image and the right-view image are set as the MV image and the AV image, respectively. Although the initial HDR images are obtained using the estimated ICRF, the DR of the initial HDR images is inherently limited by the input LDR images, which is especially noticeable in the under/over-exposed regions. Thus, to expand the DR of the MV image R_m^i, the under/over-exposed regions are detected and restored from the corresponding regions in the initial AV HDR image R_a^i.

The under/over-exposed regions are detected from I_m, as follows:

$$S(\mathbf{p}) = \begin{cases} 1, & \text{if } 0 < L_m(\mathbf{p}) < \tau_u \text{ or } \tau_o < L_m(\mathbf{p}) < 255 \\ 0, & \text{otherwise} \end{cases}, \tag{1}$$

where $S(\cdot)$ represents a binary under/over-exposed region map (1: under/over-exposed, 0: otherwise). τ_u and τ_o are the thresholds to determine the under/over-exposed regions. $\mathbf{p}$ denotes the 2D pixel coordinates and L_m denotes the luminance value of I_m. Figure 4c shows the detected under/over-exposed region map.
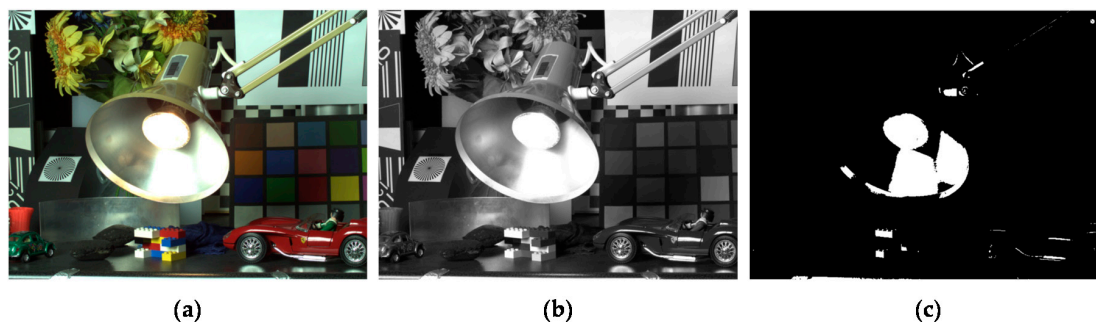
Figure 4. (**a**) Input MV LDR image, I_m; (**b**) luminance channel of I_m and L_m; and (**c**) under/over-exposed region map, S.

The corresponding under/over-exposed regions need be detected from the AV image in order to restore the radiance values of the under/over-exposed regions. For this purpose, per-pixel disparity of the AV image, D_a, is required. We estimate the disparity by combining the HVS-based cost measure and the segmentation-based aggregation method. The proposed disparity estimation method is described in detail in Section 2.2.

Unlike the conventional stereo HDR imaging methods [14–16], the segmented images are used in the proposed stereo HDR imaging method. Prior to disparity estimation, the segmented images are obtained using the two-step segmentation. To speed up the segmentation process, the initial HDR images are over-segmented into super-pixels using the super-pixel lattices (SL) method [20], as shown in Figure 5b. Since the SL method segments the image into a regular grid of super-pixels, it can be easily adapted to the graph-based algorithm. The obtained super-pixels are processed as the input pixels of the graph-based region merging (GRM) method [21]. The resultant super-pixels of the SL method are grouped into segments using the GRM to generate the segmented MV image G_m and AV image G_a, as shown in Figure 5c. These segmentation results are used for the disparity estimation and the hole-filling.
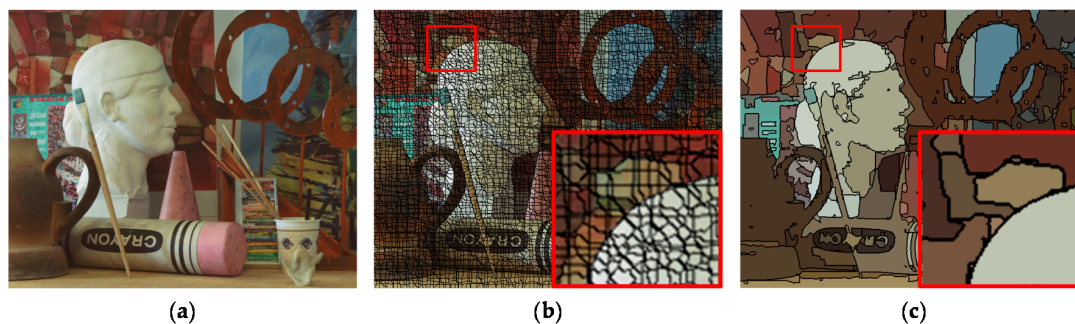

Figure 5. An example of the segmentation process: (**a**) input image; (**b**) resultant image of the super-pixel lattices [20]; and (**c**) resultant image of the graph-based region merging [21].

After obtaining the segmentation results and the AV disparity D_a, image warping is performed to align the AV HDR image with the MV HDR image. Then, the warped AV HDR image R_a^w is compensated by the hole-filling process. The proposed hole-filling method and the fusion process to reconstruct the final HDR image are presented in detail in Section 2.3.

2.2. Disparity Estimation

Prior to the estimation of the AV disparity D_a, the regions of interest in the AV image are first detected from Equation (1). To do this, the segments in G_m that include the pixels in the under/over-exposed regions are selected. The pixels in these segments are determined as interest pixels in the AV image as shown in Figure 6b. Then, each interest pixel is expanded to the horizontal direction with the maximum disparity, $d_{\max}$, as shown in Figure 6c for the estimation of the AV disparity. The proposed method only estimates the disparity of the interest regions to reduce the complexity.

(a)　　　　　　　　　　　　　　(b)　　　　　　　　　　　　　　(c)
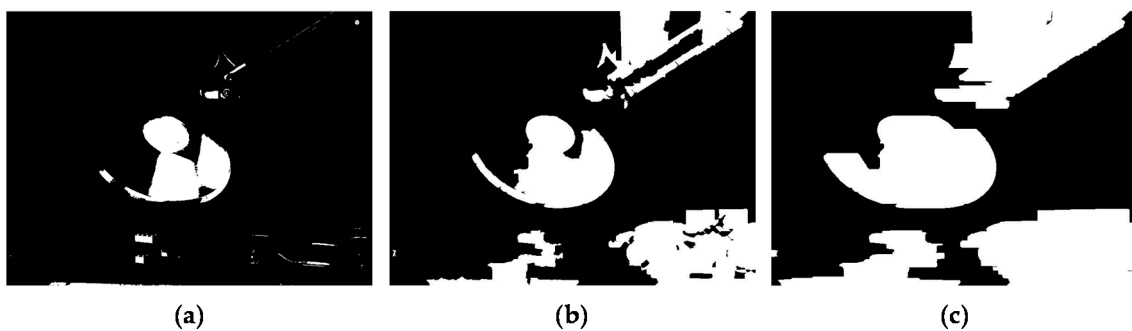
Figure 6. (**a**) Under/over-exposed region map, *S*; (**b**) a set of all the segments that contain the under/over-exposed pixels; and (**c**) interest regions.

Since it is shown that stereo matching can better perform with the HDR images than the LDR images [22], we also perform stereo matching using the initial HDR images, R_m^i and R_a^i. To compute the initial matching cost $C(\mathbf{p},d)$ at pixel $\mathbf{p}$ for the disparity value $d \in [0, d_{\max}]$, the census transform [23] is used. The performance of the census transform-based cost computation is reported to be superior to other matching cost computation methods for the images with radiometric variations [24]. In the census transform, a bit string is defined by a center pixel $\mathbf{p}$ and pixel $\mathbf{q} \in N(\mathbf{p})$, where $N(\mathbf{p})$ is a set of pixels around $\mathbf{p}$. Each bit is set to 0 if the intensity of the corresponding pixel $\mathbf{q}$ is lower than that of the pixel $\mathbf{p}$ and set to 1 otherwise. The matching cost between the pixel $\mathbf{p}$ in the AV image and a candidate pixel $\mathbf{p'}$ in the MV image is measured by computing the Hamming distance between corresponding bit strings.

In the proposed disparity estimation process, each bit of the census transform is converted into four-valued code by using a threshold, as shown in Figure 7. Moreover, the threshold for the four-valued census transform is obtained by a perceptual threshold, which is called just noticeable difference (JND). The resultant eight codes are concatenated as follows:

$$s(\mathbf{p}) = \underset{\mathbf{q} \in N(\mathbf{p})}{||} s_{\mathrm{c}}(\mathbf{p}, \mathbf{q}), \tag{2}$$

where

$$s_c(\mathbf{p}, \mathbf{q}) = \begin{cases} 0, & \text{if } R(\mathbf{q}) < R(\mathbf{p}) - T_{\mathrm{JND}}(\mathbf{p}) \\ 1, & \text{if } R(\mathbf{p}) - T_{\mathrm{JND}}(\mathbf{p}) \leq R(\mathbf{q}) < R(\mathbf{p}) \\ 2, & \text{if } R(\mathbf{p}) \leq R(\mathbf{q}) \leq R(\mathbf{p}) + T_{\mathrm{JND}}(\mathbf{p}) \\ 3, & \text{if } R(\mathbf{q}) < R(\mathbf{p}) - T_{\mathrm{JND}}(\mathbf{p}) \end{cases}, \tag{3}$$

where

$$T_{\text{JND}}(\mathbf{p}) = R(\mathbf{p}) \times 0.08, \tag{4}$$

where $\mid\mid$ denotes concatenation, $s_c(\mathbf{p},\mathbf{q})$ is the four-valued code of $\mathbf{p}$ and $\mathbf{q}$. $s(\mathbf{p})$ is the eight code string of $\mathbf{p}$, $R(\mathbf{p})$ is radiance value of p, and $T_{JND}(\mathbf{p})$ is the threshold value for p by the JND. As a result, the number of coincident codes between the code strings of radiance becomes the matching cost.
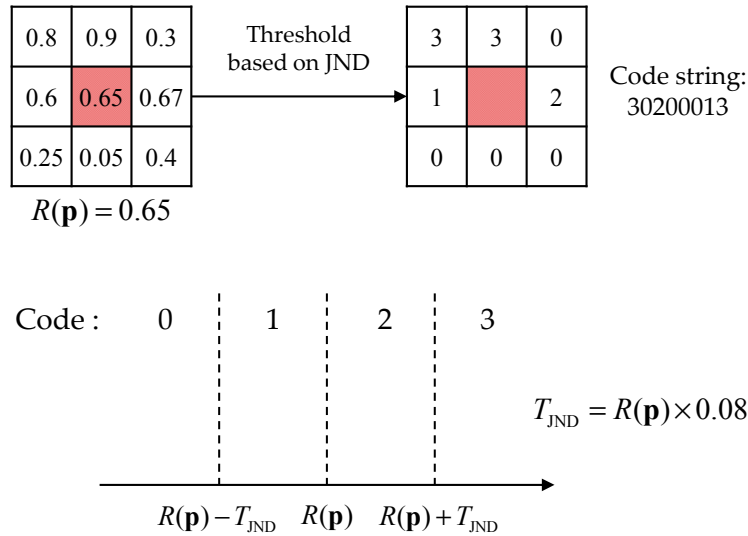
Figure 7. The process of human visual system (HVS)-based cost computation in a 3×3 local region.

Given the initial matching cost $C(\mathbf{p},d)$ at the pixel $\mathbf{p}$ for disparity value d, cost aggregation is performed using the adaptive support-weight approach [25]. The adaptive support-weights of the pixels in a given support window using the color similarity and geometric distance are used to increase the reliability of the disparity map D_a [25], as follows:

$$w_a(\mathbf{p}, \mathbf{q}) = \exp\left(-\left(\frac{\|I(\mathbf{p}) - I(\mathbf{q})\|_2}{\varepsilon_I} + \frac{\|\mathbf{p} - \mathbf{q}\|_2}{\varepsilon_S}\right)\right), \tag{5}$$

where w_a is the adaptive support-weight function and $\|\cdot\|_2$ denotes the L-2 norm. $\mathbf{p}$ and $\mathbf{q}$ denote a central pixel position and a neighboring pixel position of $\mathbf{p}$, respectively, and $I(\mathbf{p})$ represents a color components of $\mathbf{p}$. ε_I and ε_S denote predefined parameters. Since the proposed method already groups the pixels into segments, it can be assumed that the pixels in the same segment have similar disparity values [26]. Thus, the weight function w_s is defined as follows:

$$w_s(\mathbf{p}, \mathbf{q}) = \exp\left(-\frac{\|\mathbf{p} - \mathbf{q}\|_2}{\sigma^2}\right), \tag{6}$$

where σ^2 represents the variance of $w_s(\cdot)$. Then, the aggregated cost C_A is obtained by

$$C_A(\mathbf{p}, d) = \frac{\sum\limits_{\mathbf{q} \in N(\mathbf{p})} \delta(\mathbf{p}, \mathbf{q}) \times w_s(\mathbf{p}, \mathbf{q}) \times C(\mathbf{q}, d)}{\sum\limits_{\mathbf{q} \in N(\mathbf{p})} \delta(\mathbf{p}, \mathbf{q}) \times w_s(\mathbf{p}, \mathbf{q})}, \tag{7}$$

where $\delta(\cdot)$ represents an indicate function which identifies whether $\mathbf{p}$ and $\mathbf{q}$ belong to the same segment or not as follows:

$$\delta(\mathbf{p}, \mathbf{q}) = \begin{cases} 1, & \text{if } G_a(\mathbf{p}) = G_a(\mathbf{q}) \\ 0, & \text{otherwise} \end{cases}, \tag{8}$$

where G_a represents the segmented AV image [26]. In addition, to reduce the computational complexity of the cost aggregation step, cost aggregation is performed for the selected disparity candidates [27].

Finally, the WTA optimization is performed to obtain the best disparity as given below:

$$D(\mathbf{p}) = \underset{d}{\mathrm{argmax}}\, C_A(\mathbf{p}, d). \tag{9}$$

2.3. Hole-Filling for the Warped AV HDR Image

Figure 8 presents a flow chart of the warping, hole-filling, and image fusion process for reconstructing the final HDR image. Given the AV disparity D_a, the interest pixels in the AV HDR image can be aligned with the MV HDR. To do this, forward warping is performed using D_a, as follows:

$$R_a^w(\mathbf{p}') = R_a^i(\mathbf{p}) \tag{10}$$

where $\mathbf{p} = [x\ y]^\mathrm{T}$ and $\mathbf{p}' = [(x + D_a(\mathbf{p}))\ y]^\mathrm{T}$, and R_a^i and R_a^w denote the initial AV HDR image and the warped AV HDR image, respectively. However, due to inaccurate disparity values and occlusion, R_a^w involves incorrect radiance values and holes as shown in Figure 9b,c. Thus, uncertain pixels with the incorrect radiance value should be detected and removed. Then, the radiance values of holes can be restored using reliable radiance values.
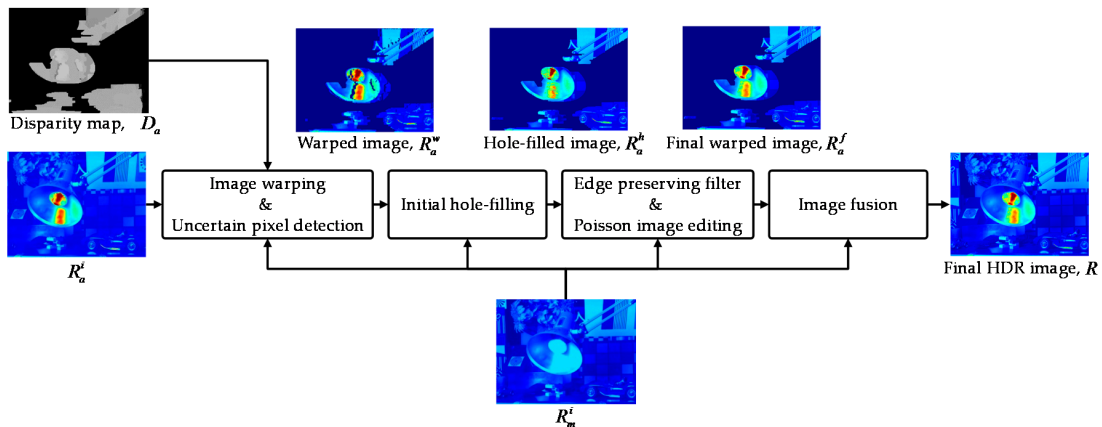

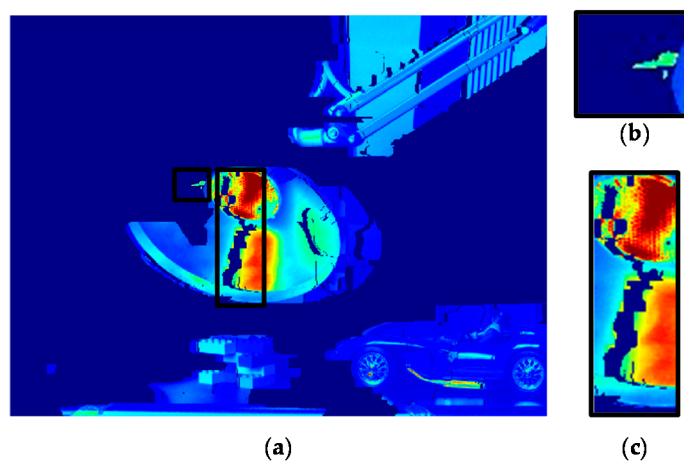**Figure 8.** A flow chart of the hole-filling and the fusion process.

Figure 9. (**a**) R_a^w; (**b**) example of incorrectly warped part enlarged with a black box in (**a**); and (**c**) example of holes enlarged with black box in (**a**).

To detect unreliable pixels, the structure of each pixel in R_a^w is compared with that of co-located pixel in the initial MV HDR image R_m^i using the structural similarity proposed in [28]. For each HDR image, the nine pixels inside a 3×3 patch centered on each pixel are converted into a bit-string by thresholding with the average of the radiance values inside the patch. The radiance value of the pixel is preserved if two bit-strings obtained from the co-located pixel in R_m^i and R_a^w are equal. Otherwise, the pixel radiance value is replaced with zero, as shown in Figure 10. In other words, the detected uncertain pixels are turned into the holes.
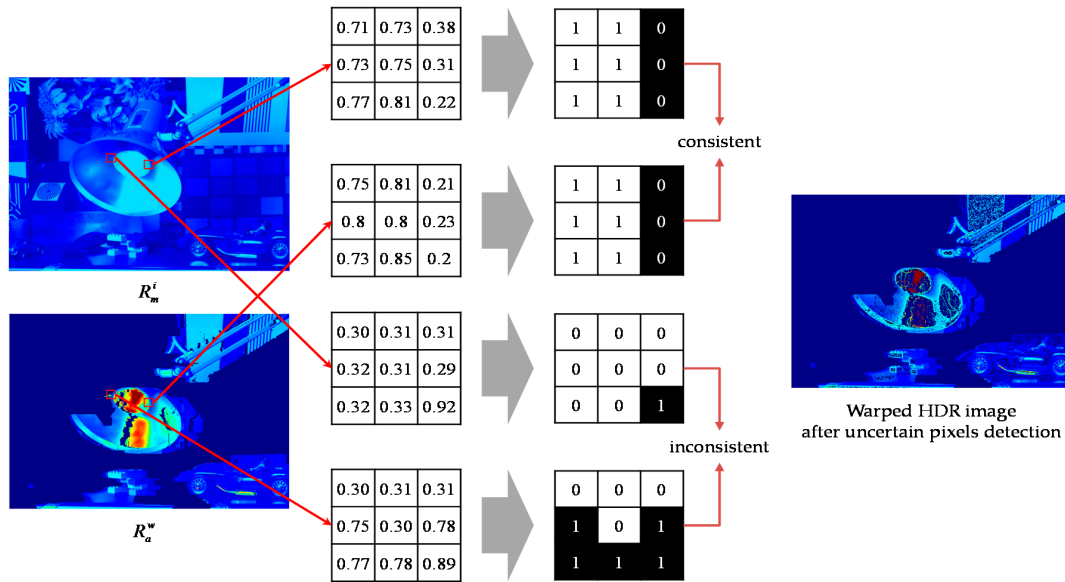

Figure 10. An example of the uncertain pixel detection.

Before proceeding to the image fusion process, it is required to restore the radiance value of the hole. To achieve this purpose, an effective hole-filling method is proposed. Let Ω represent a set of holes. To determine the restored radiance value of the pixel in hole $\mathbf{p}_h \in \Omega$, the rays along the four directions from $\mathbf{p}_h$ are first emitted. When each ray from $\mathbf{p}_h$ meets the pixel $\mathbf{q} \notin \Omega$, the radiance value of $\mathbf{q}$ is collected as the candidate radiance value of $\mathbf{p}_h$. Moreover, to collect only reliable candidate pixels, the segmented image is used. It can be assumed that the pixels in the same segment have the similar radiance values. Thus, the candidate pixel $\mathbf{q}$ is only collected inside the same segment which includes $\mathbf{p}_h$. For example, as shown in Figure 11, $\mathbf{p}_h$ has only three candidate radiance values. The radiance value of $\mathbf{p}_h$ is determined as one among the candidate values that is most similar to the radiance value of the co-located pixel in R_m^i.

Next, a filtering process is performed to improve the overall radiance. The edge-preserving filter [29] is applied for smoothing the hole-filled HDR image, R_a^h, by using the gradient of R_m^i as guidance. To further recover the texture information lost by the filtering process, Poisson image editing [30] is used as a secondary post-processing. Specifically, Poisson image editing is utilized to transfer the gradients of R_m^i to the gradients in Ω. To this end, the solution of the minimization problem is defined as follows:

$$R_a^f = \min_{R_a^e} \iint_{\Omega} \left| \nabla R_a^e - \nabla R_m^i \right|^2 \quad \text{with } R_a^e|_{\partial \Omega} = R_a^f \Big|_{\partial \Omega}, \tag{11}$$

where R_a^e represents the resultant HDR image of the edge-preserving filtering process [29], ∇ denotes the gradient operator, and $\partial \Omega$ represents the boundary of Ω. The final warped HDR image R_a^f is shown in Figure 12.
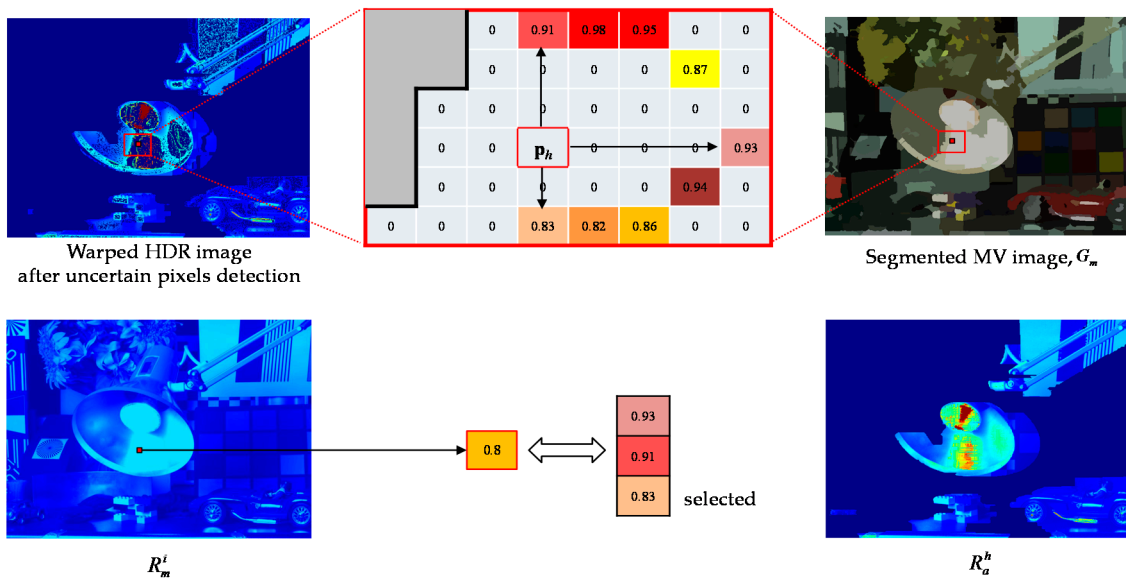
Warped HDR image
after uncertain pixels detection

Segmented MV image, G_m

R_m^i

selected

R_a^h

Figure 11. An example of initial hole-filling.

R_a^h

Edge-preserving filter
&
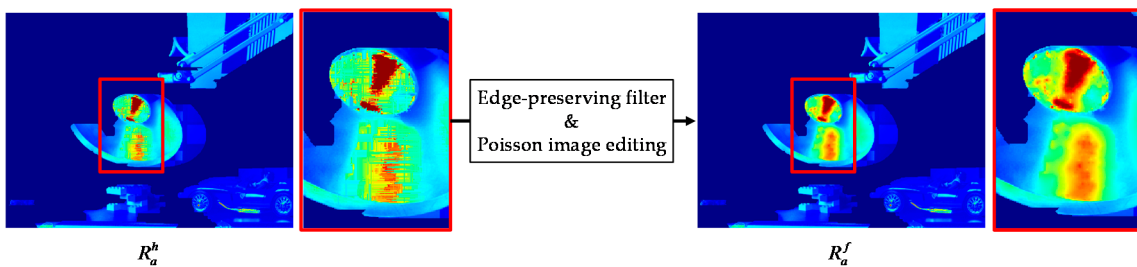Poisson image editing

R_a^f

Figure 12. Restoration radiance values for holes.

Finally, given R_a^f, the image fusion process is performed so that the radiance values of the under/over-exposed regions are determined by those of R_a^f and the radiance values of the rest regions are obtained from those of R_m^i. That is, the final HDR image R is reconstructed by fusing R_m^i with R_a^f, as follows:

$$R(\mathbf{p}) = (1 - W(\mathbf{p})) \times R_m^i(\mathbf{p}) + W(\mathbf{p}) \times R_a^f(\mathbf{p}), \tag{12}$$

where $W(\cdot)$ is the weight map for image fusion. To blend the two HDR images seamlessly around the boundary of the under/over-exposed region, the weight map W is defined by smoothing the under/over-exposed region map S using the edge-preserving filter [29], which preserves strong edges of R_m^i, as shown in Figure 13.

(a)　　　　　　　　　　　　　　　　　　**(b)**

Figure 13. (**a**) Under/over-exposed region map, S; and (**b**) weight map, W.

3. Results

3.1. Experimental Setup

To evaluate the performance of the stereo HDR imaging, the experiments were conducted on four stereo datasets: three datasets (*Aloe*, *Art*, and *Moebius*) of the Middlebury database [31] and the *IIS Jumble* dataset used in [16]. Each dataset of the Middlebury database consists of seven views (View 0–6) with three different exposure times for three illumination types (Illum 1–3). For each dataset, a chosen illumination was used for the experiments. The *Aloe* dataset has a resolution of 641 × 555 with Illum 3. The *Art* and *Moebius* datasets have a resolution 695 × 555 with Illum 2 and Illum 1, respectively. Among the seven views in the Middlebury database, View 1 and View 5 were chosen as the left-view and right-view, respectively, for the stereo HDR imaging. The *IIS Jumble* dataset is comprised of 15 different views with a resolution of 2560 × 1920. On the IIS Jumble dataset, the images from View 12 and View 13 were chosen, down-sampled by a factor of 2 in each dimension, and cropped to 800 × 600. All the experiments were conducted with the left-view images as MV images and right-view images as AV images.

All the parameters used in the proposed method were experimentally determined. To detect the under/over-exposed regions, τ_u and τ_o in Equation (1) were set to 5 and 250. For SL segmentation [20], the strip size and energy tolerance were set to 6 and 4. Furthermore, in the GRM method [21], filter variation, control value, and the minimum size of the segment were set to 0.1, 150, and 1, respectively. In the disparity estimation process, the window size for the census transform-based cost computation was set to 7 and the disparity search ranges, d_{max}, of the *Aloe*, *Art*, *Moebius*, and *IIS Jumble* were set to 100, 120, 120, and 100 pixels, respectively. The number of disparity candidates was set to 10 percent of d_{max} and σ was set to 17 in Equation (6). The spatial and range standard deviations for the edge-preserving filter [29] were, respectively, set to 20 and 0.0005, and the number of iteration steps was set to 2.

3.2. Evaluation of Performance

In order to compare the performance of the proposed stereo HDR imaging to the conventional method, the experiments were performed in two different exposure settings, the normal-long exposure and the short-long exposure. In the case of the normal-long exposure, the images with the normal exposure time and the long exposure times were used as input images, as shown in Figure 14. In the *Aloe* dataset, the exposure times of the input images are 500 ms and 2000 ms, respectively. In the *Art* and *Moebius*, the exposure times of those are 1000 ms and 4000 ms. Figure 15 shows the reference tone-mapped LDR images and the resultant tone-mapped LDR images obtained by the conventional method and the proposed method. In this paper, all the resulting HDR images were tone-mapped for visualization on the LDR display devices using the tone mapping operator [32], which was also used in the conventional method [16]. As shown in Figure 15, the superiority of the proposed method over the conventional methods [14,16] is not very convincing from the tone-mapped versions of the HDR images in the case of the normal-long exposure. To compare the objective quality of the resultant HDR images, the visual difference predictor (VDP), called HDR-VDP-2, which is a well-known image quality metric for HDR image [33], was employed. The HDR-VDP-2 ranges from 0 (worst) to 100 (best) The reference HDR images were generated using the HDR imaging method based on the bracketed images [16,34]. Table 1 shows the obtained HDR-VDP-2 scores. The proposed method achieved 2.0 and 1.3 more points than the methods of Lin et al. [14] and of Batz et al. [16] in the case of normal-long exposure, respectively. In addition, the HDR-VDP2 maps were presented for better visualization. The HDR-VDP2 maps are color-coded using the color range which represents the error values (from 0 to 100). As shown in the color bar of Figure 16, the blue and red colors represent the lowest (0) and highest (100) error values, respectively. Figure 16 shows that the differences among the conventional methods [14,16] and the proposed algorithm are marginal in the normal-long exposure case.

Figure 14. Input image pairs with normal-long exposure from the Middlebury database: (**a**) *Aloe*; (**b**) *Moebius*; and (**c**) *Art*.
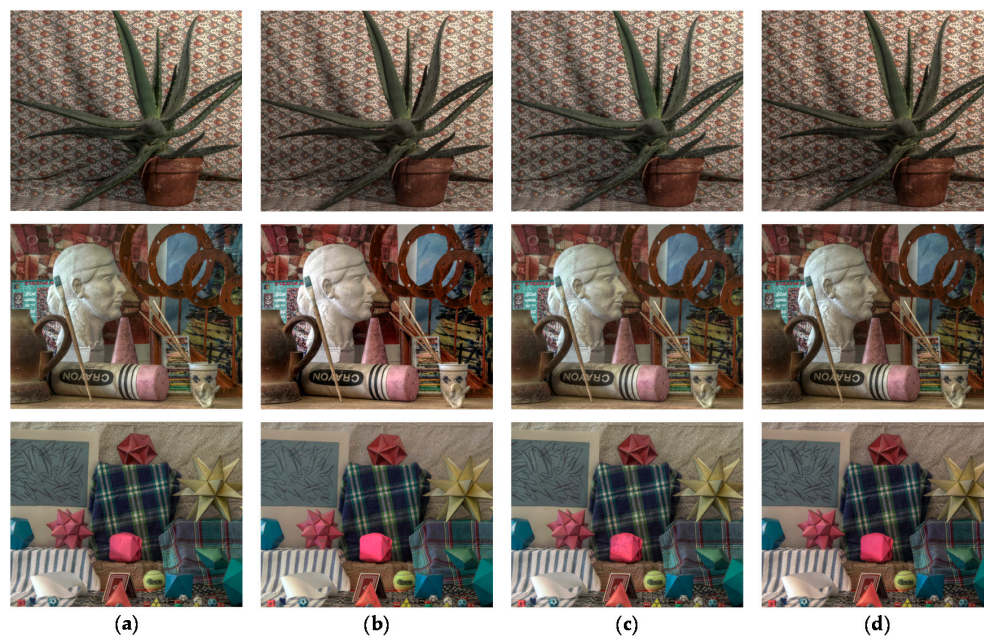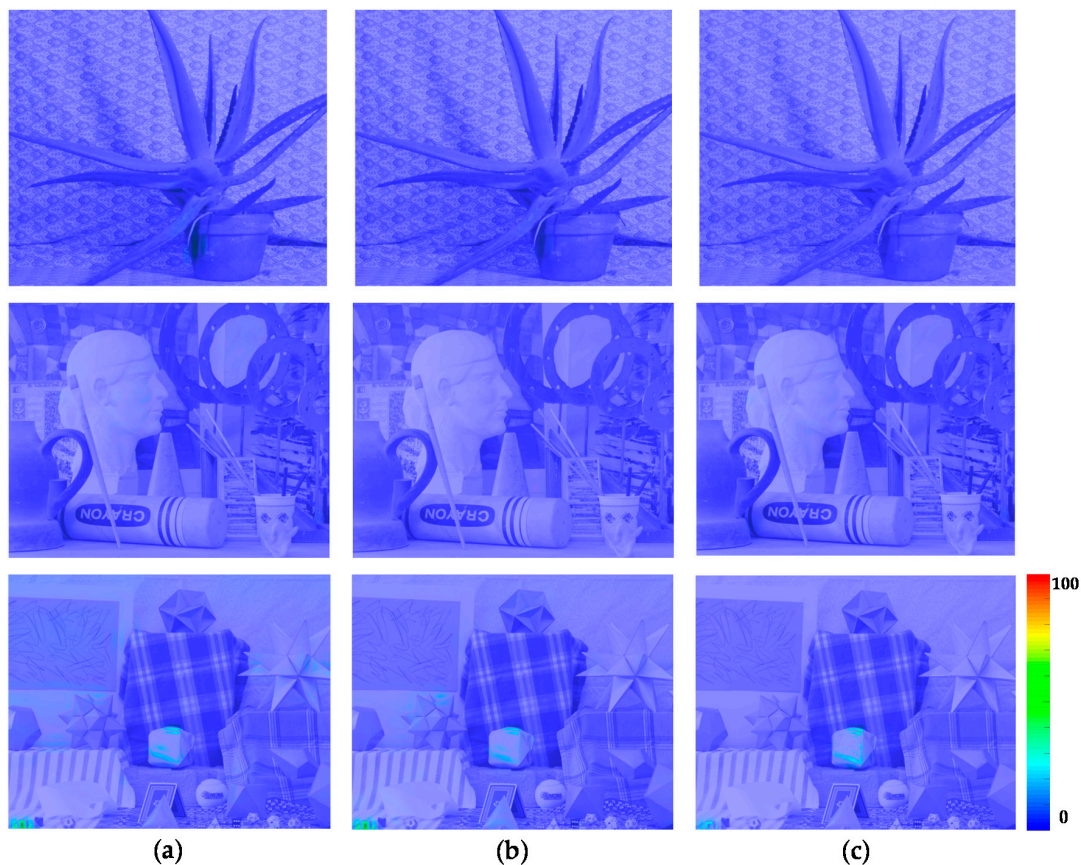

Figure 15. Resultant LDR images for the Middlebury database with normal-long exposure: (**a**) reference image; (**b**) Lin and coworkers' method [14]; (**c**) Batz and coworkers' method [16]; and (**d**) proposed method.

Table 1. Quantitative performance evaluation from the HDR-VDP-2 quality score [32].

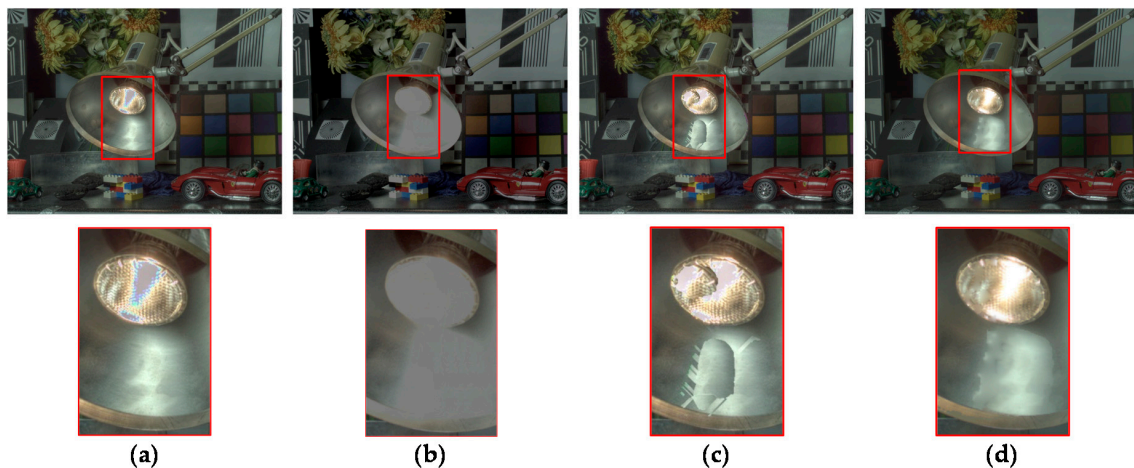| HDR-VDP-2	Normal-Long Exposure Pairs			Short-Long Exposure Pairs		
	Lin and Coworkers' Method [14]	Batz and Coworkers' Method [16]	Proposed	Lin and Coworkers' Method [14]	Batz and Coworkers' Method [16]	Proposed
Art	92.05	92.76	94.87	79.23	84.95	92.64
Aloe	91.51	92.81	94.30	78.89	83.81	90.57
Moebius	92.27	92.44	92.76	82.43	86.26	92.71
IIS Jumble				38.81	45.56	58.01
Average	91.94	92.67	**93.98**	69.84	75.15	**83.48**

Figure 16. Resultant HDR visual difference predictor 2 (HDR-VDP2) maps for the Middlebury database with normal-long exposure: (**a**) Lin and coworkers' method [14]; (**b**) Batz and coworkers' method [16]; and (**c**) proposed method.

In Figure 17, the first row and the second row represent the MV and the AV images captured by short and long exposure times, respectively. In the *Aloe* dataset, the exposure times of the input images are 125 ms and 2000 ms. In the *Art* and *Moebius*, the exposure times are 250 ms and 4000 ms, respectively. The exposure times of the *IIS Jumble* are 61 ms and 5 ms. Figure 18 shows the resultant LDR images of the *IIS Jumble* dataset in the case of short-long exposure. To highlight the differences, certain parts of the resultant images were indicated by red rectangles. As shown in Figure 18b, the Lin and coworkers' method [14] failed to reconstruct the radiance values in the over-exposed region around the light bulb. The Batz and coworkers' method [16] reconstructed the resultant HDR image with clearly visible artifacts in those regions, as shown in Figure 18c. On the other hand, the proposed method provided the reconstructed image without obvious artifacts as show in Figure 18d. Figure 19 shows the results of the conventional methods and the proposed method for the Middlebury database. In the *Aloe* dataset, as shown in Figure 19d, it seems that the Lin and coworkers' method [14] generated the HDR images without artifacts, but the method could not sufficiently restore the radiance values, as listed Table 1. The Batz and coworkers' method [16] generated artifacts at object boundaries, as shown in the Figure 19g. On the other hand, the proposed method generated the resultant image with clear object boundaries, as shown in the Figure 19j. The similar artifacts could be observed in the result of the *Art* dataset, as shown in Figure 19f,g. Lin and coworkers' method [14] made the artifacts in the right-bottom regions magnified by red rectangle. In the resultant image of Batz and coworkers' method [16], the parts of sticks of brushes and red pillars appeared repeatedly. In contrast, the proposed method reconstructed the resultant image without such artifacts, as shown in Figure 19k. In the *Moebius* dataset, while the conventional methods generated the artifacts such as those in the

magnified regions, the proposed method reconstructed the HDR image without obvious artifacts at the object boundaries, as shown in last row of Figure 19. The objective quality of the resultant HDR images of the proposed method and the conventional algorithms [14,16] in the Middlebury database and the *IIS Jumbl*e dataset are listed in Table 1. On average, the proposed method achieved a gain of 13.6 and 8.3 points in the HDR-VDP-2 quality score as compared to the methods of Lin et al. [14] and of Batz et al. [16], respectively. Figure 20 shows the HDR-VDP2 maps of the resultant images of the Middlebury database and the *IIS Jumble* dataset in the case of short-long exposure. The superiority of the proposed method over the conventional methods is clearly noticeable in the short-long exposure case as shown in Figure 20. For example, the HDR-VDP-2 maps of the *IIS Jumble* dataset obtained by the conventional methods exhibit large values around the over-exposed regions near the light bulb.
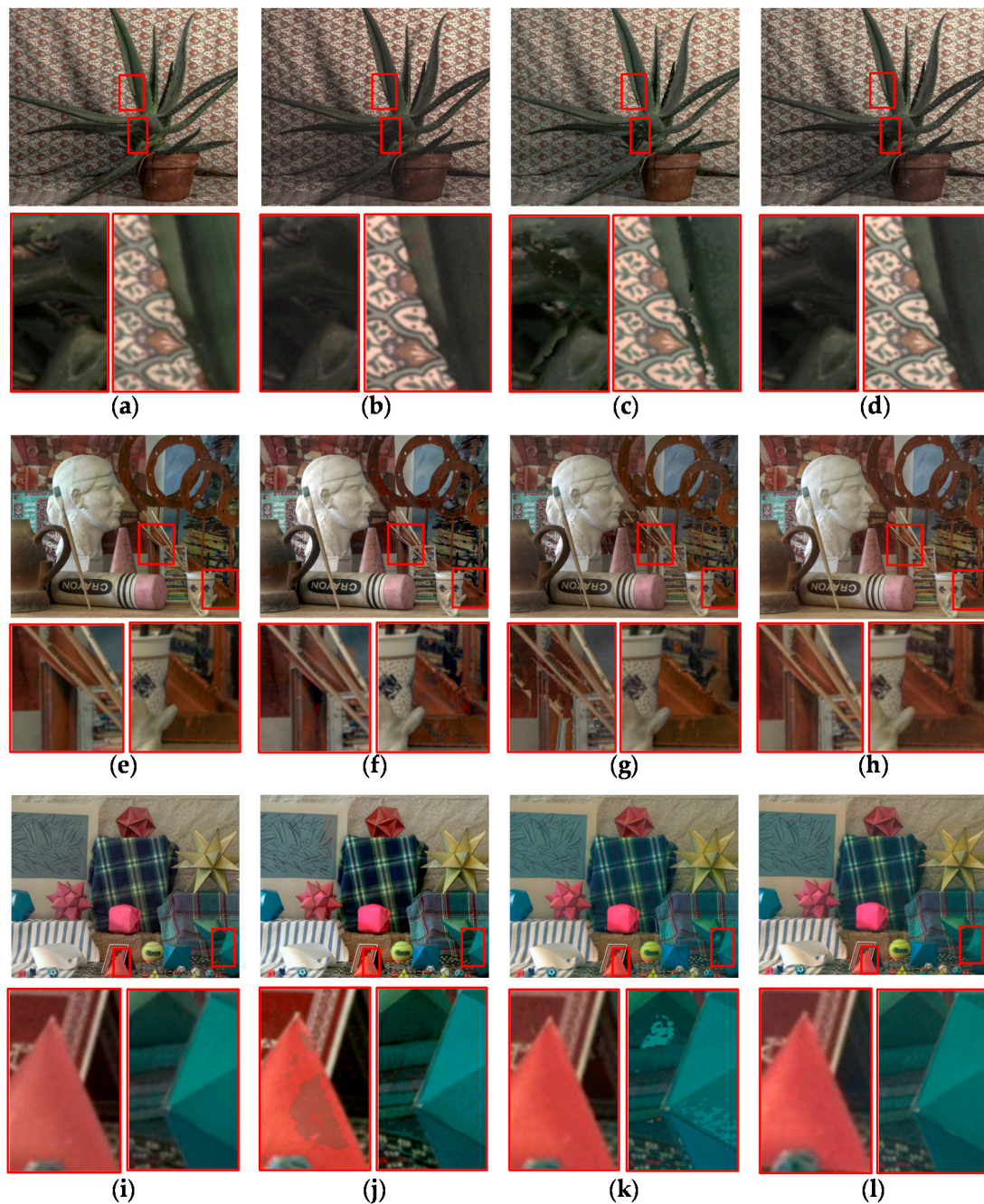

Figure 17. Input image pairs with short-long exposure: (**a**) input Middlebury images; and (**b**) input *IIS Jumble* images.

Figure 18. Resultant LDR images for the *IIS Jumble* image: (**a**) reference image; (**b**) Lin and coworkers' method [14]; (**c**) Batz and coworkers' method [16]; and (**d**) proposed method.
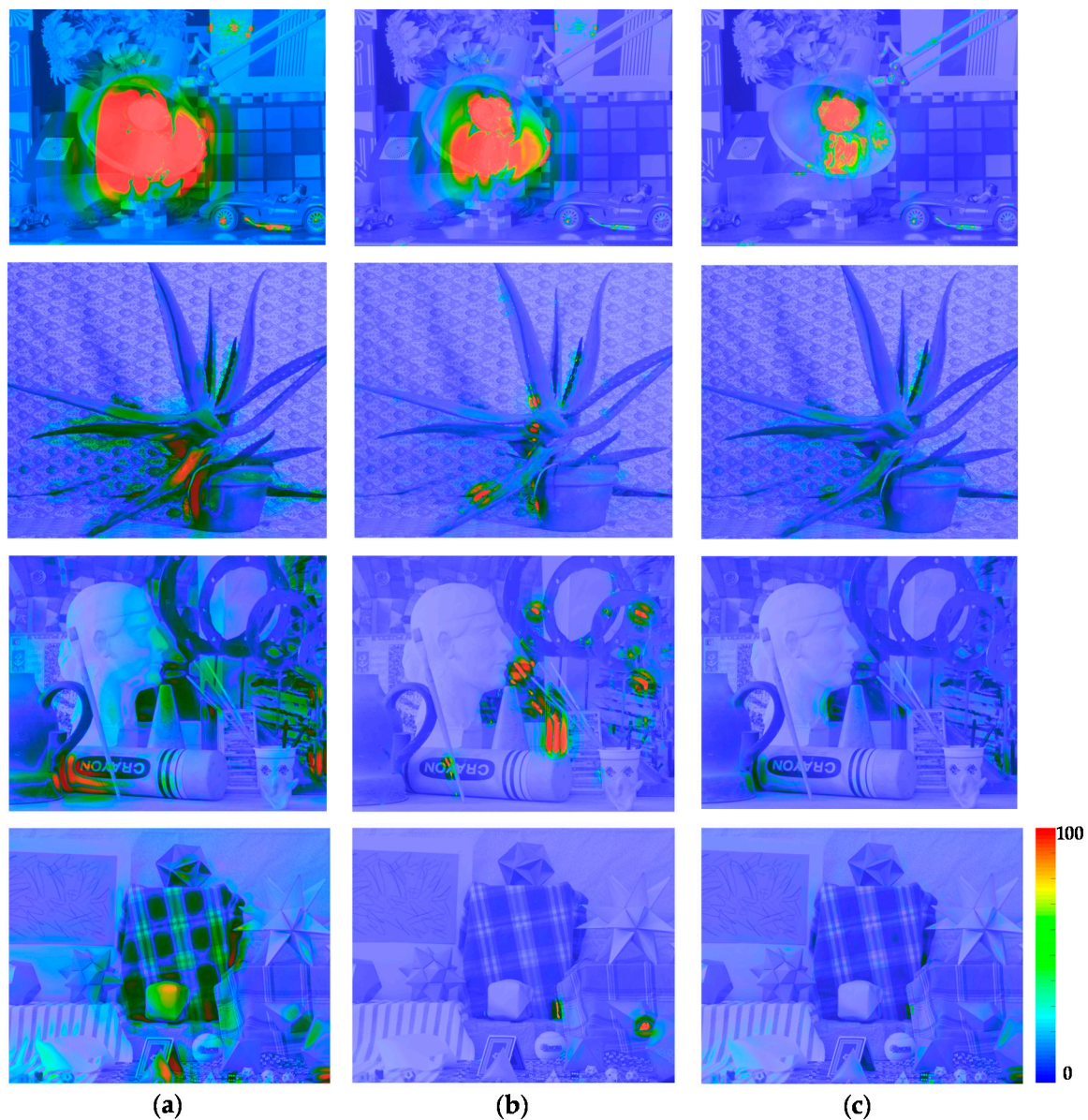
Figure 19. Resultant LDR images for the Middlebury database with short-long exposure: (**a**,**e**,**i**) reference image; (**b**,**f**,**j**) Lin and coworkers' method [14]; (**c**,**g**,**k**) Batz and coworkers' method [16]; and (**d**,**h**,**l**) proposed method.

In addition, we performed an in-depth analysis of the components of the proposed method including the ICRF estimation, rejection of the uncertain pixels, and Poisson image editing. To this end, the experiments were conducted when each method is excluded or replaced by other conventional method. Figure 21 shows the resultant images and the HDR-VDP2 maps obtained using the proposed and conventional ICRF estimation methods. The conventional method [14] generates the HDR images with clearly visible artifacts especially around object boundaries, as shown in the second row of Figure 21. The HDR-VDP2 maps clearly show the strength of the proposed method over the conventional method. Figure 22 shows that the rejection of the uncertain pixels can reduce the artifacts

resulting from inaccurate disparity values. Figure 23 shows the effects of Poisson image editing. In the proposed method, Poisson image editing step serves to further remove artifacts, as shown in Figure 23. For the quantitative performance evaluation, we measured the HDR-VDP2 scores for the sub-optimal configurations, as listed in Table 2. For notational simplicity, the rejection of the uncertain pixels and Poisson image editing are represented as rejection and PIE, respectively. It can be seen that each method is essential for reconstructing a high quality HDR image and the proposed method that includes all the methods yields the best performance.


Figure 20. Resultant HDR-VDP2 maps for the Middlebury database and the *IIS Jumble* dataset with short-long exposure: (**a**) Lin and coworkers' method [14]; (**b**) Batz and coworkers' method [16]; and (**c**) proposed method.
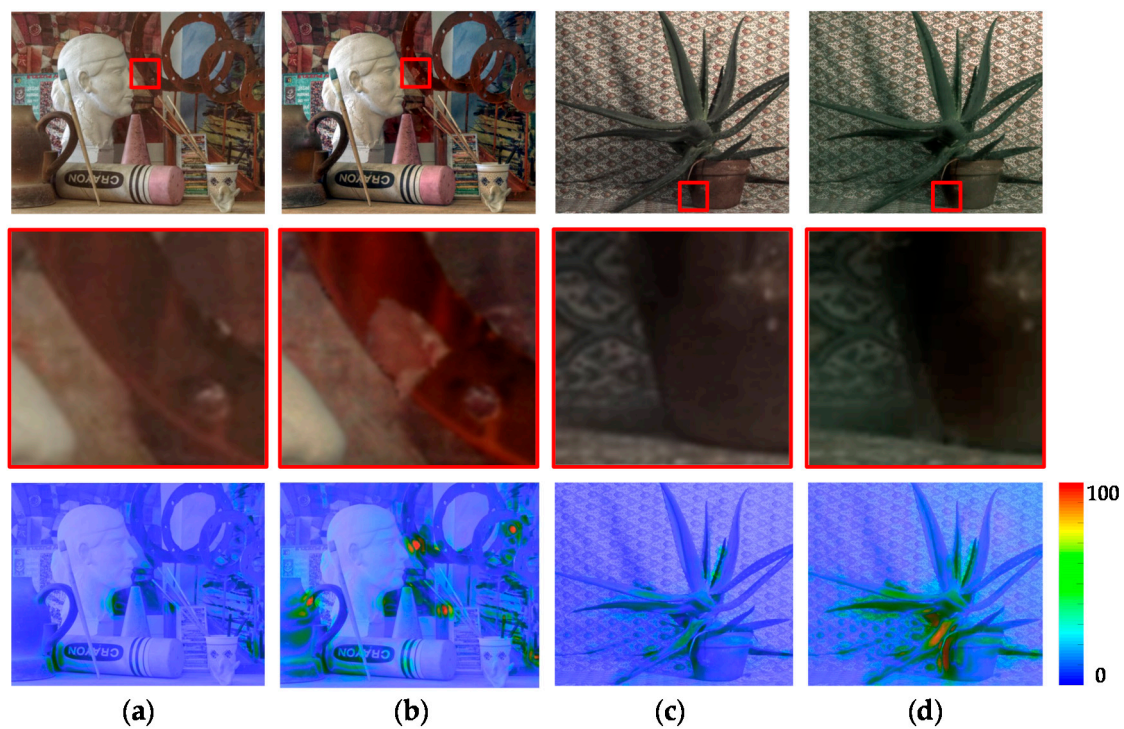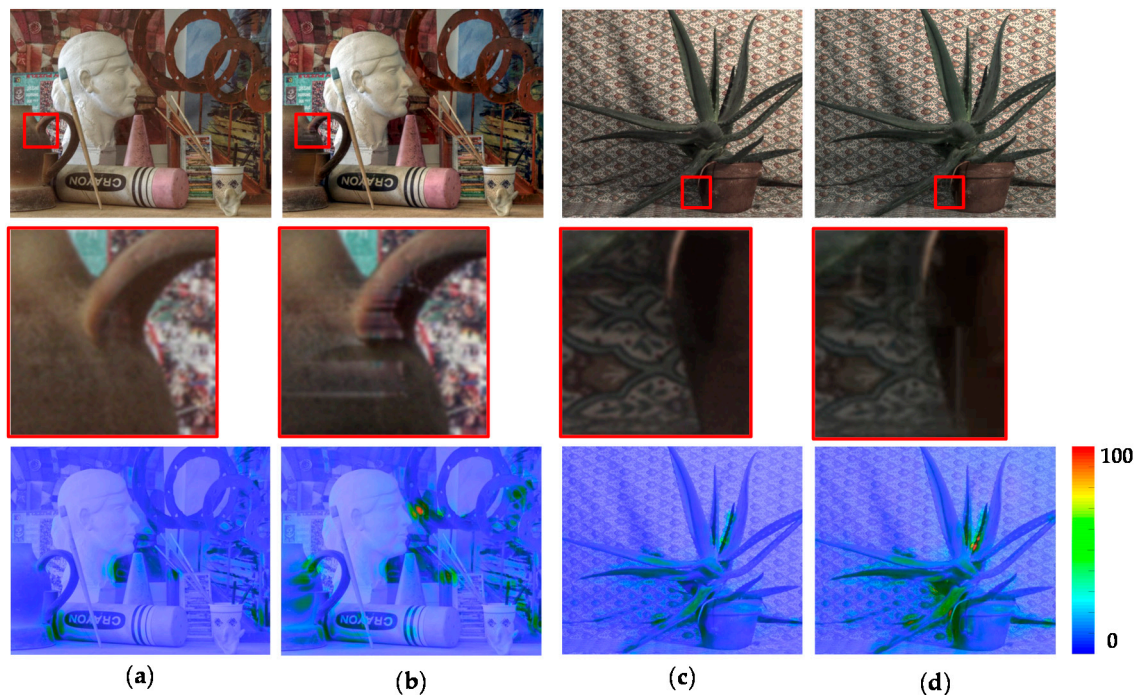
Figure 21. Resultant LDR images (first row) and their magnified sub-regions (second row) obtained with the two different ICRF estimation methods: (**a**,**c**) proposed ICRF estimation; and (**b**,**d**) conventional ICRF estimation [14]. Color-coded HDR-VDP2 maps (last row) are provided for a better performance comparison.
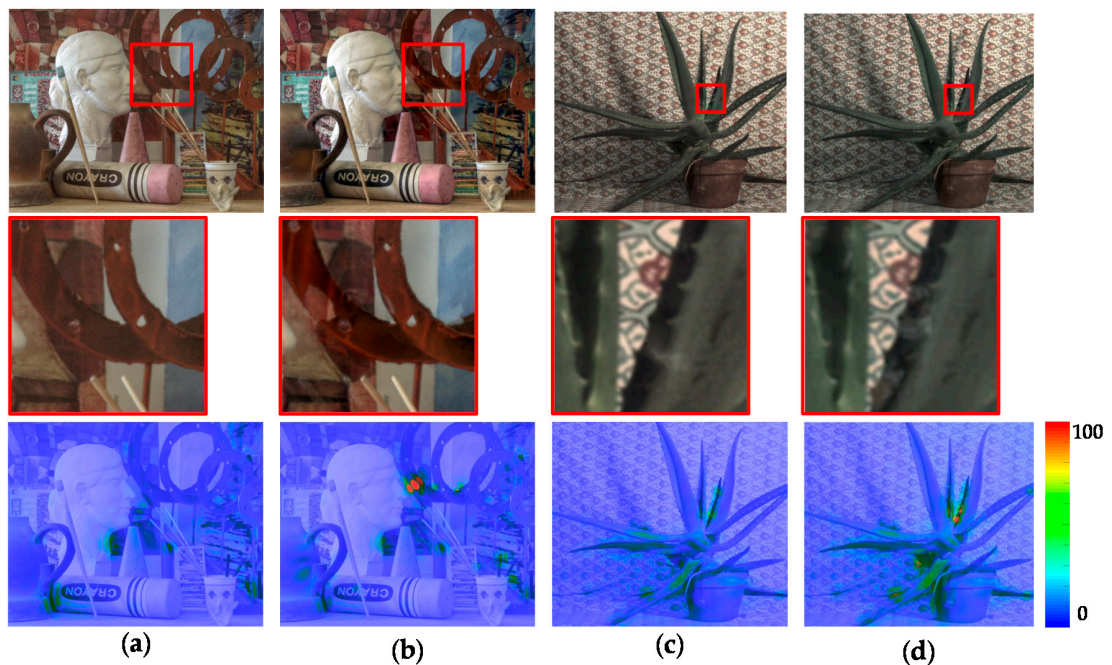
Figure 22. Resultant LDR images (first row) and their magnified sub-regions (second row) obtained with or without the proposed rejection step: (**a**,**c**) with the proposed rejection step; and (**b**,**d**) without the proposed rejection method. Color-coded HDR-VDP2 maps (last row) are provided for a better performance comparison.

Figure 23. Resultant LDR images (first row) and their magnified sub-regions (second row) obtained with or without Poisson image editing (PIE): (**a**,**c**) with PIE; and (**b**,**d**) without PIE. Color-coded HDR-VDP2 maps (last row) are provided for a better performance comparison.

Table 2. In-depth quantitative performance evaluation of the proposed method with different configurations.

HDR-VDP-2	*Art*	*Aloe*	*Moebius*	*IIS Jumble*
Proposed	92.64	90.57	92.71	58.01
Proposed with conventional ICRF	86.61	81.69	87.25	52.95
Proposed without Rejection	84.91	86.25	84.56	41.93
Proposed without PIE	86.44	86.66	87.78	49.57
Proposed without Rejection & PIE	82.48	85.81	82.63	36.22
Proposed with conventional ICRF & without PIE	82.26	78.18	85.90	49.88
Proposed with conventional ICRF & without Rejection	81.34	78.35	84.68	38.84
Proposed with conventional ICRF & without Rejection and PIE	78.17	77.95	81.87	35.96

4. Conclusions

In this paper, the method to reconstruct an HDR image was presented using stereo LDR images with different exposure times. Since the HDR image is reconstructed from images simultaneously captured with different exposure, the stereo HDR imaging method has a merit of being relatively free from the ghosting artifact problem in comparison with the HDR imaging method based on the temporal exposure bracketing. However, the performance of the stereo HDR imaging depends on the following processes: ICRF estimation to obtain the initial HDR images, disparity estimation to align the input images, and image warping followed by image fusion to reconstruct a high-quality HDR image. Unlike the conventional stereo HDR imaging methods, the proposed method mainly improved two major sub-processes: (1) disparity estimation; and (2) image warping followed by image fusion. In the disparity estimation process, the disparity was estimated only at the pixels in the interest regions detected using the segmented image. Moreover, the HVS-based cost computation and segmentation-based cost aggregation were proposed to accurately estimate the disparity. In the image warping and fusion process, effective hole-filling was performed to enhance the warped HDR image. Then, the final HDR image was reconstructed through edge-preserving filter-based image

fusion. The experimental results demonstrated the superiority of the proposed stereo HDR imaging method compared to the conventional method.

Author Contributions: Won-Jae Park contributed to the design and experimental verification of the proposed method as well as the realization of the paper. Seo-Won Ji and Seok-Jae Kang contributed to the execution of the experiments. Seung-Won Jung participated in the discussion about the proposed method and contributed to the analysis of the results. Sung-Jea Ko also participated in the discussion about the proposed method and contributed to the idea development and the paper revision.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shafie, S.; Kawahito, S.; Itoh, S. A dynamic range expansion technique for CMOS image sensors with dual charge storage in a pixel and multiple sampling. *Sensors* **2008**, *8*, 1915–1926. [CrossRef] [PubMed]
2. Shafie, S.; Kawahito, S.; Halin, I.A.; Hasan, W.Z.W. Non-linearity in wide dynamic range CMOS image sensors utilizing a partial charge transfer technique. *Sensors* **2009**, *9*, 9452–9467. [CrossRef] [PubMed]
3. Martínez-Sánchez, A.; Fernández, C.; Halin, I.A.; Navarro, P.J.; Iborra, A. A Novel Method to Increase LinLog CMOS Sensors' Performance in High Dynamic Range Scenarios. *Sensors* **2011**, *11*, 8412–8429. [CrossRef] [PubMed]
4. Mertens, T.; Kautz, J.; Van Reeth, F. Exposure fusion: A simple and practical alternative to high dynamic range photography. *Comput. Graph. Forum* **2009**, *28*, 161–171. [CrossRef]
5. Gallo, O.; Gelfandz, N.; Chen, W.-C.; Tico, M.; Pulli, K. Artifact-free high dynamic range imaging. In Proceedings of the IEEE International Conference on Computational Photography, San Francisco, CA, USA, 16–17 April 2009; pp. 1–7.
6. Li, S.; Kang, X. Fast multi-exposure image fusion with median filter and recursive filter. *IEEE Trans. Consum. Electron.* **2012**, *58*, 626–632. [CrossRef]
7. Zhang, W.; Cham, W.-K. Gradient-directed multiexposure composition. *IEEE Trans. Image Process.* **2012**, *21*, 2318–2323. [CrossRef] [PubMed]
8. Jacobs, K.; Loscos, C.; Ward, G. Automatic high dynamic range image generation for dynamic scenes. *IEEE Comput. Graph. Appl.* **2008**, *28*, 84–93. [CrossRef] [PubMed]
9. Sen, P.; Kalantari, N.K.; Yaesoubi, M.; Darabi, S.; Goldman, D.B.; Shechtman, E. Robust patch-based HDR reconstruction of dynamic scenes. *ACM Trans. Graph.* **2012**, *31*, 203. [CrossRef]
10. Hu, J.; Gallo, O.; Pulli, K.; Sun, X. HDR Deghosting: How to Deal with Saturation? In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 1163–1170.
11. Lu, J.; Yang, H.; Min, D.; Do, M.N. Patch Match Filter: Efficient edge-aware filtering meets randomized search for fast correspondence field estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 1854–1861.
12. Nayar, S.K.; Mitsunaga, T. High dynamic range imaging: Spatially varying pixel exposures. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head Island, SC, USA, 15 June 2000; pp. 472–479.
13. Selmanovic, E.; Debattista, K.; Bashford-Rogers, T.; Chalmers, A. Generating stereoscopic HDR images using HDR-LDR image pairs. *ACM Trans. Appl. Perception.* **2013**, *10*, 3. [CrossRef]
14. Lin, H.-Y.; Chang, W.-Z. High dynamic range imaging for stereoscopic scene representation. In Proceedings of the 16th IEEE International Conference on Image Processing, Cairo, Egypt, 7–10 November 2009; pp. 4305–4308.
15. Sun, N.; Mansour, H.; Ward, R. HDR image construction from multi-exposed stereo LDR images. In Proceeding of the 17th IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010; pp. 2973–2976.
16. Batz, M.; Richter, T.; Garbas, J.-U.; Papst, A.; Seiler, J.; Kaup, A. High dynamic range video reconstruction from a stereo camera setup. *Signal Process. Image Commun.* **2014**, *29*, 191–202. [CrossRef]

17. Tursun, O.T.; Akyuz, A.O.; Erdem, A.; Erdem, E. The State of the Art in HDR Deghosting: A Survey and Evaluation. *Comput. Graph. Forum* **2015**, *34*, 683–707. [CrossRef]

18. Fusiello, A.; Trucco, E.; Verri, A. A compact algorithm for rectification of stereo pairs. *Mach. Vis. Appl.* **2000**, *12*, 16–22. [CrossRef]

19. Debevec, P.E.; Malik, J. Recovering high dynamic range radiance maps from photographs. In Proceedings of the ACM SIGGRAPH 2008 classes, Los Angeles, CA, USA, 11–15 August 2008; p. 31.

20. Moore, A.P.; Prince, S.J.D.; Warrell, J.; Mohammed, U.; Jones, G. Superpixel lattices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.

21. Felzenszwalb, P.F.; Huttenlocher, D.P. Efficient graph-based image segmentation. *Int. J. Comput. Vis.* **2004**, *59*, 167–181. [CrossRef]

22. Akhavan, T.; Yoo, H.; Gelautz, M. Evaluation of LDR, tone mapped and HDR stereo matching using cost-volume filtering approach. In Proceeding of the European Signal Processing Conference, Lisbon, Portugal, 1–5 September 2014; pp. 1617–1621.

23. Zabih, R.; Woodfill, J. Non-parametric local transforms for computing visual correspondence. In Proceedings of the European Conference on Computer Vision, Stockholm, Sweden, 2–6 May 1994; pp. 151–158.

24. Hirschmuller, H.; Scharstein, D. Evaluation of stereo matching costs on images with radiometric differences. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 1582–1599. [CrossRef] [PubMed]

25. Yoon, K.-J.; Kweon, I.S. Adaptive support-weight approach for correspondence search. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 650–656. [CrossRef] [PubMed]

26. Tombari, F.; Mattoccia, S.; Stefano, L.D. Segmentation-Based Adaptive Support for Accurate Stereo Correspondence. In Proceedings of the Pacific-Rim Symposium on Image and Video Technology, Santiago, Chile, 17–19 December 2007; pp. 427–438.

27. Min, D.; Lu, J.; Do, M.N. Joint histogram-based cost aggregation for stereo matching. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2539–2545. [PubMed]

28. Kang, S.-J.; Lee, D.-H.; Ji, S.-W.; Kim, C.-S.; Ko, S.-J. A novel method to generate the ghost-free wide dynamic range image. In Proceedings of the IEEE International Conference on Consumer Electronics, Las Vegas, NV, USA, 7–11 January 2016; pp. 97–98.

29. Gastal, E.S.L.; Oliveira, M.M. Domain transform for edge-aware image and video processing. In Proceedings of the ACM SIGGRAPH 2011, Vancouver, BC, Canada, 7–11 Auguest 2011; p. 69.

30. Perez, P.; Gangnet, M.; Blake, A. Poisson image editing. In Proceedings of the ACM SIGGRAPH 2003, San Diego, CA, USA, 27–31 July 2003; pp. 313–318.

31. Middlebury Stereo Vision Page. Available online: http://vision.middlebury.edu/stereo/ (accessed on 21 June 2017).

32. Mantiuk, R.; Myszkowski, K.; Seidel, H.-P. A perceptual framework for contrast processing of high dynamic range images. In Proceedings of the Second Symposium on Applied Perception in Graphics and Visualization, La Coruña, Spain, 26–28 August 2005; pp. 87–94.

33. Mantiuk, R.; Kim, K.J.; Rempel, A.G.; Heidrich, W. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. In Proceedings of the ACM SIGGRAPH 2011, Vancouver, BC, Canada, 7–11 Auguest 2011; p. 40.

34. Robertson, M.; Borman, S.; Stevenson, R. Dynamic range improvement through multiple exposures. In Proceedings of the International Conference on Image Processing, Kobe, Japan, 24–28 October 1999; pp. 159–163.