

Super High Dynamic Range Imaging

Takehito Hayami*, Masayuki Tanaka*, Masatoshi Okutomi*, Takashi Shibata[†] and Shuji Senda[†]

*Department of Mechanical and Control Engineering
Tokyo Institute of Technology

[†]NEC Corporation

Abstract—We propose a novel high dynamic range (HDR) imaging algorithm for the scenes that contain an extremely wide range of scene radiance. In the HDR imaging, several images are taken under different exposures. Those images usually have displacement from one another due to camera and/or object motions. The challenge of the super HDR imaging is to align those images because any image contains "lost" regions where texture information is completely lost due to overexposure or underexposure. We propose an image alignment algorithm based on similarities of region shapes instead of the similarities of the textures. Experimental comparisons demonstrate that the proposed algorithm outperforms state-of-the-art algorithms.

I. INTRODUCTION

A typical consumer's digital camera has limited dynamic range for luminance. One cannot simultaneously capture images with both extremely bright and extremely dark regions with a single shot. Therefore, computational high dynamic range (HDR) imaging has been studied. Classical HDR imaging algorithms [1], [2], [3] first estimate a camera response function with multiple images taken under different exposures. Then, scene radiances are obtained from the pixel values using the camera response function. For these processes, it is assumed that the scene is static and that the images are taken from a fixed viewpoint. However, if we apply the classical HDR imaging algorithms to the dynamic scene or images taken using a hand-held camera, the resultant HDR image suffers from ghost artifacts.

To avoid the ghost artifacts, two kinds of methods are adopted. The first is to build a specialized HDR camera system that includes multiple cameras and beam splitters [4], [5] so that images can be simultaneously taken under different exposures without displacement. However, such systems are bulky and expensive.

The second is to develop algorithms to reduce the ghost artifacts using images with displacement. These algorithms are further classifiable into two approaches. The first approach is a weighting approach, which assigns lower weight to ghost pixels or dynamic regions to reduce the ghost artifacts [6], [7], [8], [9], where the weight is used to compose the HDR image. Heo et al. [10] detected the dynamic region based on a joint probability density function and a Gaussian weight. The joint probability density function is estimated with similarities based on textures. The second one is an optical-flow-based approach [11], [12], [13], [14], [15], [16]. In this approach, a reference image is first determined among different exposure images. Then, the optical flow of each image with respect

to the reference image is estimated for aligning the images. Similarities based on textures are used to estimate the optical flow. Finally, the HDR image is composed with the aligned images using a naive HDR composing algorithm.

These two approaches commonly depend on similarities based on textures. In other words, texture information is assumed to be obtainable from all different exposure images. However, some textures are completely lost because of the limited dynamic range for each image compared with the range of the scene radiance. The existing algorithms that rely on textures fail to align the images because there is no texture information in some regions. Consequently, severe artifacts appear in the composed HDR images, if the range of the scene radiance becomes wide.

For such a scene with an extremely wide range of scene radiance, we propose a novel HDR imaging algorithm, which we call super HDR imaging. A key idea of the proposed algorithm is to align images based on the shapes of segmented regions instead of textures, so that we can align extremely underexposed or overexposed regions. Then, we collect the regions with appropriate exposures based on the alignment and compose the HDR image. We apply our proposed HDR imaging algorithm to actual scenes with an extremely wide range of scene radiance. Experimental comparisons demonstrate that the proposed algorithm outperforms existing state-of-the-art algorithms.

II. CHALLENGES OF SUPER HDR IMAGING

To discuss challenges of the super HDR imaging, we consider a one-dimensional image as shown in Fig. 1. Fig. 1(a) shows scene radiance where the horizontal and vertical axes respectively represent position and radiance. This scene includes textures at extremely bright and extremely dark regions. If we take this scene under long exposure, then the texture information in the dark region can be captured. However, texture information in the extremely bright region is lost because of saturation. On the contrary, in the image taken under short exposure, the texture information in the bright region can be retained without the saturation, but the texture information in the dark region is blacked out or extremely degraded by noise. In addition, displacement usually occurs between images because of the camera and/or scene motion. Schematics of those relations are shown in Fig. 1(b). Even after adjusting pixel values to scene radiances, it is a challenging problem to align extremely dark or extremely bright regions where the texture information is lost. As described in

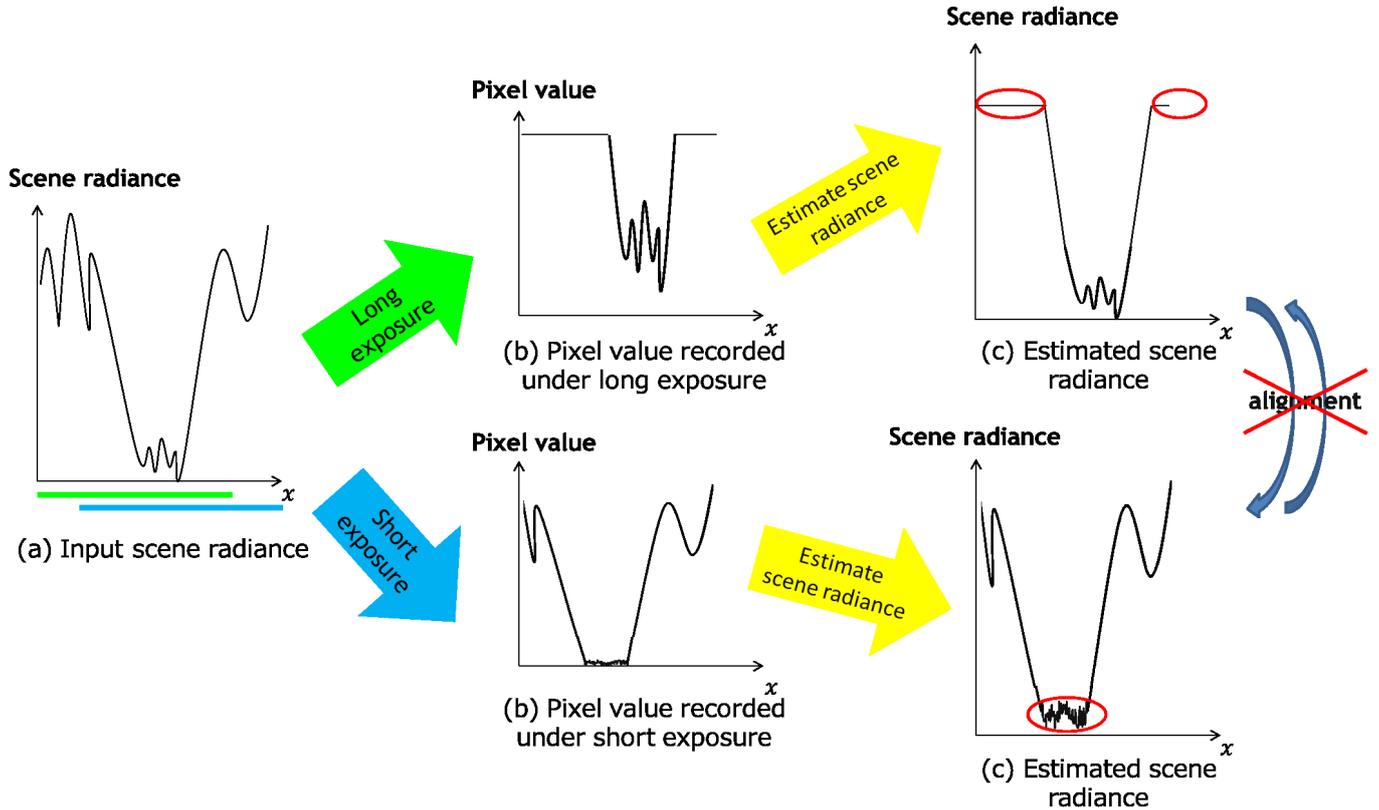


Fig. 1. We consider a one-dimensional scene: (a) is the input scene radiance. If we take this scene under short exposure and long exposure, then (b) images are captured. In addition, displacement usually occurs between images because of the camera or scene motion. Even if these images are converted to scene radiance, they cannot be mutually aligned.

the previous section, existing de-ghosting algorithms depend on similarities based on texture information. Therefore, the existing de-ghosting algorithms cannot be applied to such a scene with regions where texture information is completely lost.

As described in details in the following sections, we propose an algorithm to align those regions based on the segmented region shapes. Fig. 2 shows an overview of the proposed algorithm. First, we perform segmentation based on the luminance for each exposure image. Then, we align images by comparing the shape of the segmented region to that in the reference image. For the alignment, we also use texture information where possible. Once alignment is done for all regions, we can compose the HDR image using the naive compositing algorithm.

III. PROPOSED ALGORITHM

In the proposed algorithm, each exposure image is segmented based on luminance into three regions: saturated, appropriate, and blacked-out. We refer this segmentation process to trinarization. The main idea of the proposed algorithm is to compose the HDR image by collecting the appropriate regions from all exposure images.

Let $\{I_1, I_2, \dots, I_N\}$ be images taken under different exposures, where images are sorted in descending order of

exposure. The proposed algorithm consists of three steps: trinarization, region alignment, and merging.

A. Trinarization by luminance

The camera response function is assumed to be known, so pixel values of images with different exposure can be converted to scene radiance. We explain how the saturated, the appropriate, and the blacked-out ranges are determined. The observable maximum scene radiance, Z_{\max} , is defined by the scene radiance corresponding to the maximum value of the image intensity. This observable maximum scene radiance depends on the exposure. We define the saturated range of each exposure by the range τZ_{\max} to Z_{\max} , where $\tau = 0.90$ is used for our experiments. To define the appropriate and the blacked-out ranges, we consider two different exposures as shown in Fig. 3: short and long exposures. Let Z_{\max}^s and Z_{\max}^l respectively denote the observable maximum scene radiance of the short and long exposures. The ratio of the short exposure to the long exposure is known and denoted by r . We set the lower limit of the appropriate range for the short exposure as $\tau r Z_{\max}^s$, so that it corresponds to the upper limit of the appropriate range for the long exposure, τZ_{\max}^l . Consequently, the range from $\tau r Z_{\max}^s$ to τZ_{\max}^l is defined as the appropriate range. And the range below the appropriate range is defined as blacked-out range, in which pixels should be replaced with those of the long exposure in the following steps. These definitions are

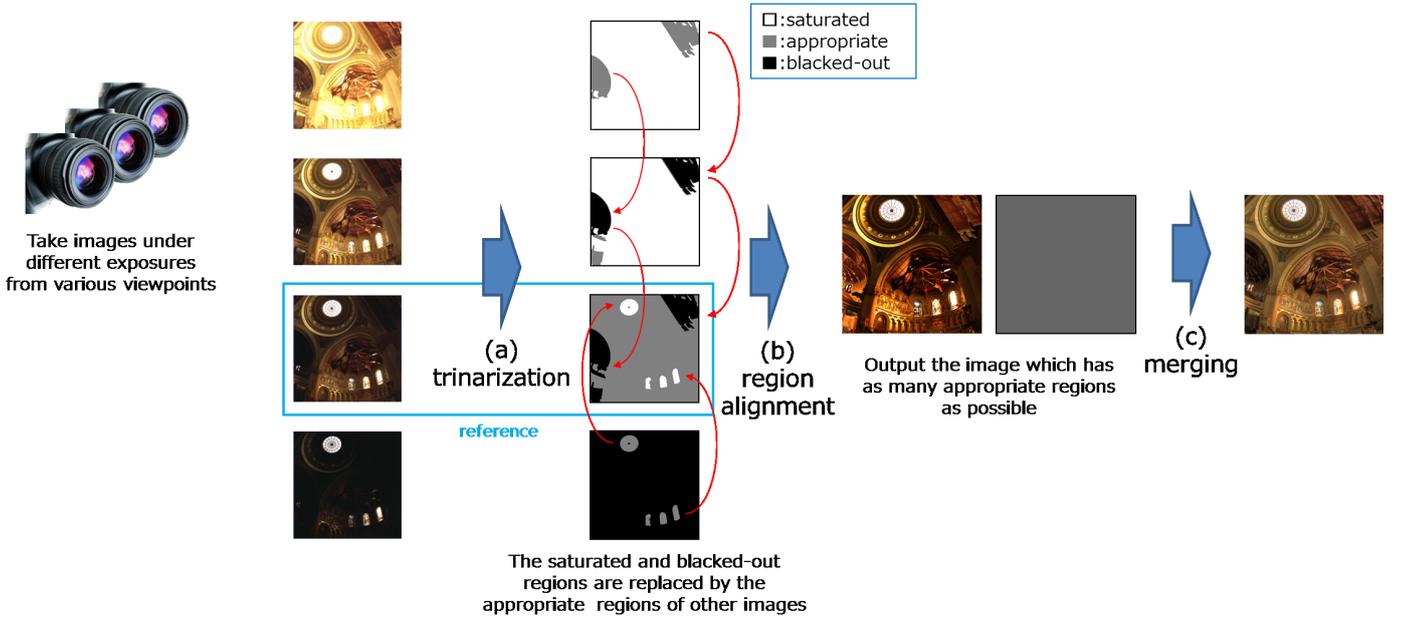


Fig. 2. Here, (a), (b), and (c) are overviews of our algorithm. (a) All input images are trinarized by luminance into three regions: saturated, appropriate, and blacked-out. Then, the image with the largest area of the appropriate region is defined as the reference image. (b) Images are aligned based on the shapes of segmented regions. (c) The HDR image is composed by merging aligned images.

applied for any two consecutive exposures. In the proposed algorithm, each exposure image is trinarized based on the above definition for each range and spacial smoothness for the segments. We use the Graph-Cut optimization algorithm [17] for this trinarization.

B. Alignment based on the shapes of segmented regions

For the image alignment, we first select a reference image. The image with the largest area of the appropriate region is selected as the reference image R . Now, we consider replacing the blacked-out regions of the reference image with the appropriate regions of the longer exposure images. Here, we consider that the image I_k is selected as the reference image. Then, the next long exposure image I_{k-1} is the first source image. Thanks to the trinarization described in the previous section, the blacked-out region of the reference image can be expected to be associated with a combined region of appropriate and blacked-out regions of the source image. Therefore, we replace the blacked-out regions of the reference image with the appropriate and blacked-out regions of the source image, by aligning each region based on the shape of the segmented regions. The reference image updated by this replacement may still have blacked-out regions in the source image. Then, the same process should be applied using I_{k-2} as the source image.

We repeat these processes until all longer exposure images are merged. In the same manner, we replace the saturated regions of the reference image by the appropriate and saturated regions of next short exposure image I_{k+1} as the source image. This process is also repeated until all shorter exposure images are merged.

Now, we describe our alignment method based on the

shapes of the segmented regions. We first convert the reference and source trinarized images to binary images. For the reference image, the appropriate regions are set as 1 and the blacked-out and the saturated regions are set as 0. For longer exposure source images, the appropriate and the blacked-out regions are set as 1. For a shorter exposure source image, the appropriate and the saturated regions are set as 0 and the blacked-out regions are set as 1. These conversions enable us to compare the shapes of the regions as a jigsaw puzzle. Then, the alignment can be conducted by minimizing the cost function as

$$E_{\text{shape}}(\mathbf{p}) = \frac{1}{|\Omega'|} \sum_{\mathbf{x} \in \Omega'} \tilde{R}(\mathbf{x}) \oplus \tilde{S}(\mathbf{x} + \mathbf{p}), \quad (1)$$

where \tilde{R} and \tilde{S} are the binary image generated by binarizing the trinary images of the reference image R and source image S , \mathbf{x} is the position of the pixel, \mathbf{p} represents the displacement, \oplus represents exclusive OR operator and Ω' is the dilated region of the current region of the reference image, Ω . We set the two thresholds for the trinarization as shown in Fig. 3, so that the blacked-out region of the short exposure image corresponds to the appropriate and the blacked-out regions of the long exposure image and that the saturated region of the long exposure image corresponds to the appropriate and the saturated regions of the short exposure image. Therefore, we convert the trinarized images to binarized images. If the texture information remains there, then we add the similarity cost based on the texture information as

$$E_{\Omega}(\mathbf{p}) = E_{\text{shape}}(\mathbf{p}) + \alpha E_{\text{texture}}(\mathbf{p}). \quad (2)$$

Therein, $E_{\text{texture}}(\mathbf{p})$ is weighted-SSD expressed as

$$E_{\text{texture}}(\mathbf{p}) = \frac{1}{|\Omega|} \sum_{\mathbf{x} \in \Omega} w(R(\mathbf{x})) (R_{\text{sr}}(\mathbf{x}) - S_{\text{sr}}(\mathbf{x} + \mathbf{p}))^2, \quad (3)$$

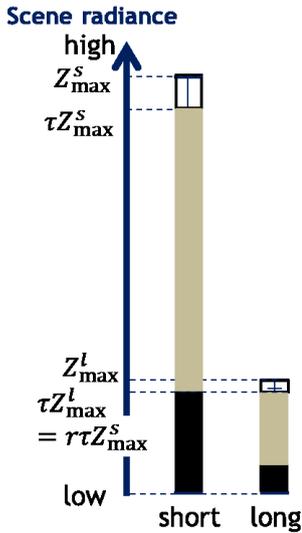


Fig. 3. τZ_{\max} is the threshold between saturated regions and appropriate regions for each image. $\tau r Z_{\max}$ is the threshold between appropriate regions and blacked-out regions for each image. We trinarize all images based on these thresholds.

where R_{sr} and S_{sr} are generated by converting R and S to scene radiance. In addition, $w(i)$ is a weight used as a soft-threshold to down-weight underexposed and overexposed regions. In general, we can align regions sufficiently by minimizing only E_{shape} . We use the texture information as additional information. Although detailed comparisons are presented in Section 4, the texture energy term does not affect the results to any great degree. Consequently, difference p is calculated for all reference regions Ω .

C. Merging

Once we align the images, we can simply apply the naive HDR algorithm to compose the HDR image. In addition, we apply a blending technique over boundaries to alleviate the brightness change if the correct camera response function is not obtained.

IV. EXPERIMENTAL RESULTS

For the evaluation, we take images at different viewpoints under different exposures. First, we compose the HDR images with alignments based on the three different cost functions: texture only, shape only, and texture and shape. The cost function of texture only corresponds to the existing optical flow approach. Fig. 4 presents results of each cost function. The left is the result obtained by minimizing only E_{texture} in eq. (2). As described in section II, it is difficult to use texture similarity when a reference image has an overexposed or an underexposed part. Misalignments of the saturated regions cause artifacts as shown in Fig. 4(a). Fig. 4(b) and 4(c) show similar results, where α is 0 and 1.5×10^{-5} in eq. (2). The shape cost function mainly contributes to the alignment.

We compare the result of our algorithm with state-of-the-art algorithms. Input images are taken at different positions. These camera positions are presented in Fig. 5. The distance

position1	position2	position3
position4	position5	position6
position7	position8	position9

Fig. 5. Camera position. The distance between two adjacent cameras either horizontally or vertically is 30 mm.

between two adjacent cameras aligned vertically or horizontally is 30 mm. Fig. 6 presents input images that are taken at different exposure times. The exposure times are 1 ms, 4 ms, 16 ms, 64 ms, 256 ms, and 1024 ms from the left image. Fig. 7 shows our results and those of existing algorithms. We applied the photomatrix [18] tone mapping for our proposed algorithm and Sen's algorithm [16], while Heo's algorithm [10] includes tone mapping in itself. Comparison of these results shows that our algorithm can correctly reconstruct an overexposed region of the reference image. Severe artifacts exist in those regions in the results of other existing algorithms. Note that this scene includes the characters of "HDR" in the lighting (the brightest region) and can be visible in our result. The exposure ratio of this scene is more than one thousand. The proposed algorithm can reconstruct those scenes with the super wide range of scene radiance.

Fig. 9 shows another result. The input images are shown in Fig. 8. Exposure times of input images are 4 ms, 16 ms, 64 ms, 256 ms, and 1024 ms from the left. Our resultant HDR image appears good for the entire image regions including the most brightest part, where the characters on the label of the whiskey bottle can be visible.

V. CONCLUSIONS

We have proposed a novel algorithm to produce the HDR image with images taken of the scene including an extremely wide range of intensity levels under different exposures. As shown in the experimental results, it is difficult for the existing algorithms to produce the HDR image with those images because all of those images include overexposed and/or underexposed regions. At such regions, textures are completely lost and the alignment based on the texture similarity dose not work. In our proposed algorithm, each image is segmented by luminance so that the segmented regions fit to the corresponding regions in the other images. By aligning the regions based on their shapes, even the regions in which the texture is completely lost can be aligned. Consequently, our algorithm can produce the super HDR image for the scene with an extremely wide range of scene radiance.

REFERENCES

- [1] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *ACM SIGGRAPH 2008 classes*. ACM, 2008, p. 31.

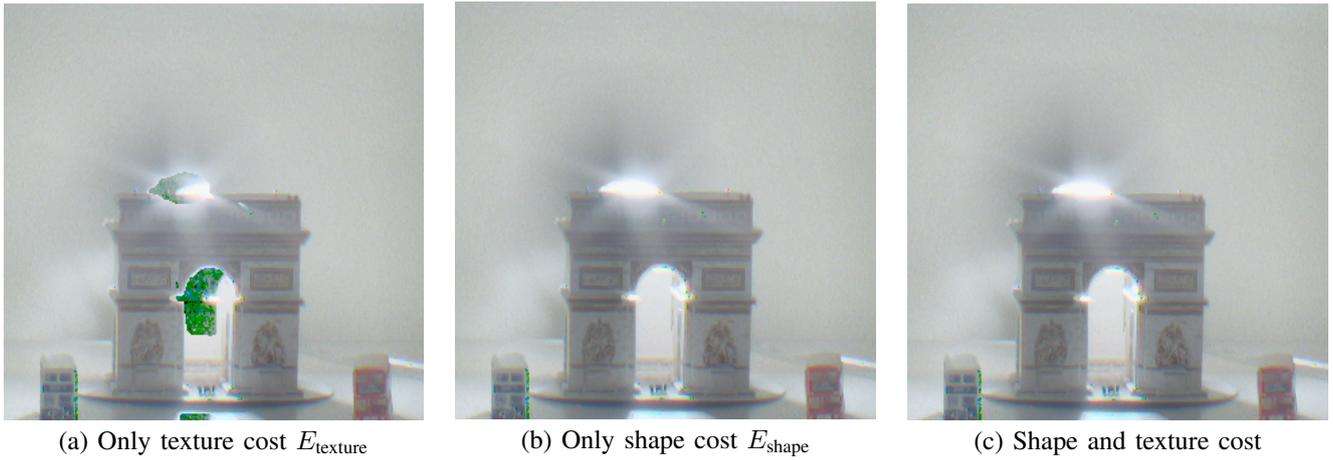


Fig. 4. These results are generated from three input images taken at different positions. These HDR images are tone-mapped with MATLAB tone-mapping. The left image is obtained by minimizing only E_{texture} in our algorithm. The result shows how the mistake of correspondence between regions arises. Furthermore, the middle shows the result obtained by minimizing only E_{shape} . The right shows that obtained by minimizing the cost function E_{Ω} (eq. (2)). These results are similar.

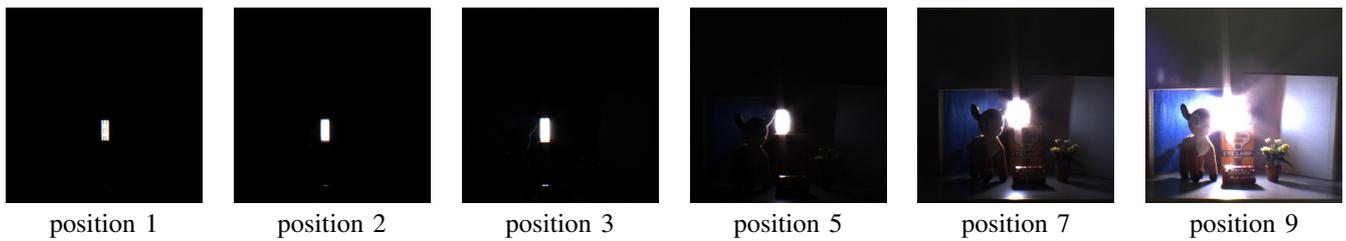


Fig. 6. These position numbers correspond to Fig. 5. Exposure times are 1 ms, 4 ms, 16 ms, 64 ms, 256 ms, and 1024 ms from the left image.



Fig. 7. At left is our result. The middle image shows the result presented by Heo et al. [10]. At right is the result of Sen et al. [16]. The lighting part is extremely bright compared with others. The result of Heo et al. has big errors in the middle part due to the misalignment. The result of Sen et al. cannot correctly reconstruct the part of light, either. In our result, the vertical word "HDR" can be correctly visible on the LED light.



Fig. 8. These images were taken at positions 1, 3, 5, 7, and 9. Exposure times were 4 ms, 16 ms, 64 ms, 256 ms, and 1024 ms from the left image.



Fig. 9. At left is our result. In the middle is the result presented by Heo et al. [10]. At the right is the result of Sen et al. [16]. In the results of existing methods the whiskey bottle on the right is severely degraded. The part of the whiskey bottle is too bright, so that texture similarity is difficult to use. Our result, which is based on shape similarity, shows the shape of the bottle and the words on the label clearly.

- [2] Mann, Picard, S. Mann, and R. W. Picard, "On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures," in *Proceedings of IS and T*, 1995, pp. 442–448.
- [3] T. Mitsunaga and S. Nayar, "Radiometric Self Calibration," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, Jun 1999, pp. 374–380.
- [4] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen, "A versatile hdr video production system," in *ACM Transactions on Graphics (TOG)*, vol. 30, no. 4. ACM, 2011, p. 41.
- [5] S. Nayar and T. Mitsunaga, "High Dynamic Range Imaging: Spatially Varying Pixel Exposures," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, Jun 2000, pp. 472–479.
- [6] W. Zhang and W.-K. Cham, "Gradient-directed multiexposure composition," *Image Processing, IEEE Transactions on*, vol. 21, no. 4, pp. 2318–2323, 2012.
- [7] K. Jacobs, C. Loscos, and G. Ward, "Automatic high-dynamic range image generation for dynamic scenes," *Computer Graphics and Applications, IEEE*, vol. 28, no. 2, pp. 84–93, 2008.
- [8] O. Gallo, N. Gelfand, W. Chen, M. Tico, and K. Pulli, "Artifact-free high dynamic range imaging," *IEEE International Conference on Computational Photography (ICCP)*, April 2009.
- [9] E. A. Khan, A. O. Akyuz, and E. Reinhard, "Ghost removal in high dynamic range images," in *Image Processing, 2006 IEEE International Conference on*. IEEE, 2006, pp. 2005–2008.
- [10] Y. S. Heo, K. M. Lee, S. U. Lee, Y. Moon, and J. Cha, "Ghost-free high dynamic range imaging," in *Computer Vision—ACCV 2010*. Springer, 2011, pp. 486–500.
- [11] S. Mangiat and J. Gibson, "High dynamic range video with ghost removal," in *Proc. of SPIE Vol*, vol. 7798, 2010, pp. 779812–1.
- [12] R. M. Anna Tomaszewska, "Image Registration for Multi-exposure High Dynamic Range Image Acquisition," *Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)*, 2007.
- [13] C. Aguerrebere, J. Delon, Y. Gousseau, and P. Musé, "Simultaneous HDR image reconstruction and denoising for dynamic scenes," in *IEEE International Conference on Computational Photography*, 2013, pp. 31–41.
- [14] J. Hu, O. Gallo, K. Pulli, and X. Sun, "Hdr deghosting: How to deal with saturation ?" in *CVPR*, 2013.
- [15] M. Gupta, D. Iso, and S. Nayar, "Fibonacci Exposure Bracketing for High Dynamic Range Imaging," in *IEEE International Conference on Computer Vision (ICCV)*, Dec 2013, pp. 1–8.
- [16] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman, "Robust Patch-Based HDR Reconstruction of Dynamic Scenes," *ACM Transactions on Graphics (TOG) (Proceedings of SIGGRAPH Asia 2012)*, vol. 31, no. 6, pp. 203:1–203:11, 2012.
- [17] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 11, pp. 1222–1239, 2001.
- [18] "Photomatix pro 2012, commercially-available hdr processing software." [Online]. Available: <http://www.hdrsoft.com/>.