# FEATURE BASED GHOST REMOVAL IN HIGH DYNAMIC RANGE IMAGING

Hwan-Soon Sung[1], Rae-Hong Park[1], Dong-Kyu Lee[1], and SoonKeun Chang[2]

[1]Department of Electronic Engineering, School of Engineering, Sogang University,
35 Baekbeom-ro (Sinsu-dong), Mapo-gu, Seoul 121-742, Korea
[2]Samsung Electronics Co., Ltd., Suwon, Gyeonggi-do 443-742, Korea

## ABSTRACT

This paper presents a technique to reduce the ghost artifacts in a high dynamic range (HDR) image. In HDR imaging, we need to detect the motion between multiple exposure images of the same scene in order to prevent the ghost artifacts. First, we establish correspondences between the aligned reference image and the other exposure images using the zero-mean normalized cross correlation (ZNCC). Then, we find object motion regions using adaptive local thresholding of ZNCC feature maps and motion map clustering. In this process, we focus on finding accurate motion regions and on reducing false detection in order to minimize the side effects as well. Through experiments with several sets of low dynamic range images captured with different exposures, we show that the proposed method can remove the ghost artifacts better than existing methods.

## 1. INTRODUCTION

Most sensors can detect only a restricted range of light intensity values. In a captured image, we cannot observe what we see in a real-world scene. More specifically, in a low dynamic range (LDR) image, a pixel with very low irradiance will appear as noisy or saturated black pixel, whereas that with very high irradiance as saturated white pixel containing incorrect information of the scene. Faithfully representing the information of the real-world scene is a challenging task. General approach to solve this problem is to produce a high dynamic range (HDR) image by combining a set of LDR images (LDRIs) captured with different exposure settings [1–4], which is so-called HDR imaging. Each LDR image provides partial information of radiance values of the scene. Combining irradiance values from multiple LDRIs, an HDR image (HDRI) can cover a wide range of radiance values that the real-world scene has. Ideally, irradiance value at one pixel is supposed to be consistent with those at corresponding pixels of the same scene taken with different exposure settings. However, if LDRIs are not aligned, inconsistencies can occur between corresponding pixels of the same scene taken at different exposure settings, which makes an HDRI suffer from ghost artifacts [4, 5].

This paper focuses on local motion detection of objects in HDR imaging. Generally, global motion comes from shaking a hand-held camera during capturing a set of LDRIs [6, 7], whereas local motion is caused by moving objects in a scene during shooting [5]. Since it takes comparatively long time to capture multi-exposure images, we cannot assure that pixels at the same location in different exposure images represent the same radiance information of the real-world scene in this situation. If the area of moving objects is relatively large in a scene, estimation accuracy of a camera response function can be affected by unaligned or misaligned part of images

since the possibility that samples for the camera response function are taken from the moving objects gets higher.

In this paper, we propose a feature based ghost removal method in HDR imaging. The rest of the paper is structured as follows. Sections 2 and 3 present previous works on ghost removal and conventional HDRI generation methods, respectively. Section 4 proposes a framework for ghost removal consisting of saturated region separation, feature map generation, adaptive local thresholding, motion map clustering, and radiance map generation. Experimental results are given in Section 5. First, we discuss the image capture setting such as the number of LDRIs and exposure value (EV) spacing which affect the quality of the HDRI [8]. Then, we compare performance of the proposed ghost removal method with existing methods. Experiments with various test image sets show that the proposed method can remove the ghost artifacts effectively. Finally, conclusions are given in Section 6.

## 2. PREVIOUS WORKS

Kang *et al*. estimated the displacement between pixels in a moving object in differently exposed images, using conventional motion estimation techniques such as optical flow method [6]. But motion estimation methods do not work reliably with differently exposed images because change of illumination gives poor motion estimation results and also they may fail if a non-rigid object is deformed. Especially, it is hard to find correspondences when occlusions or saturated regions exist in LDRIs. Therefore, we can say that a motion estimation based ghost removal method has a clear limitation on obtaining a high-quality HDRI.

Khan *et al*. proposed a method to remove ghost artifacts in an HDRI without object detection and motion estimation [9]. The key idea is to reduce the contribution of the pixels in the non-static part of the scene in constructing the final radiance map. The result of the algorithm seems quite promising. However, the algorithm is slow due to iteration process and shows poor results if the scene does not predominantly represent a static background. Pedone and Heikkila [10] extended Khan *et al*.'s method by introducing the weight propagation scheme which propagates the low weight of the object pixels to the surrounding pixels within the same object. Gallo *et al*. divided each of the LDRIs into the patches and then measured the consistencies of irradiance values between the corresponding patches [11]. They assumed that the irradiance at a given pixel remains the same unless the scene changes. One of disadvantages of the methods described above is that the overall performance of the algorithms can be affected by misalignment and small errors in camera response function estimation because they find the object motion in the radiance domain after camera response function estimation.

Jinno and Okuda used a Markov random field (MRF) model to estimate the displacement between two LDRIs [12]. Jacobs *et al*. proposed an entropy-based method [13]. They assumed that the local entropy does not change between differently exposed images. So they computed the entropy of each LDRI and then compare the difference of entropies with a pre-defined threshold to find the motion regions. In addition, because they found the object motion in the intensity domain before estimation of camera response function, the performance of motion detection is independent of the estimation accuracy of camera response function. Min *et al*. proposed a histogram based ghost removal method [14]. They used multi-level threshold maps, a modified version of the median threshold bitmap (MTB), to detect the local motion regions. This method is simple and fast, however a large portion of false detection exists, causing artifacts in an HDRI.

On the other hand, a single image based HDR imaging method [15] produces multiple LDRIs from a single image using local histogram stretching. Since it uses a single input image, it inherently does not need to remove ghost artifacts from global and local motion during exposure

time. However, there is a limit to cover a wide range of radiance values when compared to other HDR imaging methods using LDRIs captured with different exposure.

## 3. HDRI GENERATION

HDRI reconstruction begins with estimation of camera response function, which is a non-linear function that converts irradiance values to image intensity values [16]. Once the camera response function is estimated, we can obtain the irradiance map for each image by converting intensity values to relative irradiance values. The log-irradiance of the image can be written as

$$\ln E_j(x, y) = g(Z_j(x, y)) - \ln \Delta t_j \qquad (1)$$

where $E_j(x, y)$ denotes the irradiance at $(x, y)$ in $j$-th image and $g$ represents the estimated inverse camera response function. $Z_j(x, y)$ is the intensity value at $(x, y)$ in $j$-th image and $\Delta t_j$ denotes the exposure time of $j$-th image. Ideally, the irradiances of the corresponding pixels over multiple images with different exposure settings should be the same except for the pixels located in moving area or non-stationary regions.

However, due to nonlinearity introduced by quantization process, which transforms the irradiance value to the intensity value, and errors in estimation of camera response function, the recovered irradiance can be slightly different from the actual radiance values. Additionally, the recovered irradiances consist of the finite number of values since $g(Z_j(x, y))$ can take only discrete values due to quantization of $Z_j(x, y)$. Inaccurate estimation of irradiance values causes degradation in the signal to noise ratio (SNR) in a final HDRI. We need as many irradiance values at each pixel as possible to increase the SNR of the final HDRI.

## 4. PROPOSED GHOST REMOVAL METHOD

In this section, we propose a ghost-free HDRI generation method. Figure 1 shows the overall block diagram of the proposed method. Let $J$ denote the total number of LDRIs. We define input LDRI set as $\mathbf{Z} = \{Z_1(x, y), Z_2(x, y), \ldots, Z_J(x, y)\}$ with in order of increasing exposure, where $(x, y)$ is the two-dimensional (2-D) pixel coordinates. For example, $Z_J(x, y)$ is the LDRI with the longest exposure time among the input LDRI set $\mathbf{Z}$.

First, we select a reference image among LDRIs. The reference image is selected so that it does not contain the objects that we do not want to have in an HDRI [11]. Or simply select a mid-exposure image in the input LDRI set $\mathbf{Z}$, because in general a mid-exposure image contains the least number of over-/under-saturated regions. Then, we use a scale-invariant feature transform (SIFT) based alignment method [17] that finds translation as well as rotation of a camera in order to generate aligned LDRIs $\tilde{Z}_j(x, y)$, where tilde represents the image alignment. Next, we find motions using luminance images $\tilde{L}_j(x, y)$ of each $\tilde{Z}_j(x, y)$ $(j \neq R)$ [2], where $R$ represents index of the reference image. Pixels with $M_j''(x, y) = 1$, $j \neq R$, give locations of the pixels corrupted by object motion in $\tilde{Z}_j(x, y)$. After weight maps $W_j(x, y)$ and weighting factors $ME_j(x, y)$ for ghost removal are calculated for each aligned LDRI, a final radiance map $\hat{E}(x, y)$ is generated by a weighted average of irradiance maps in the logarithm domain, where all aligned LDRIs are transformed into irradiance maps with inverse camera response function $g(\cdot)$. The irradiance values of the pixels in the final motion maps with $M_j''(x, y)$, $j \neq R$, will be disregarded during this process.
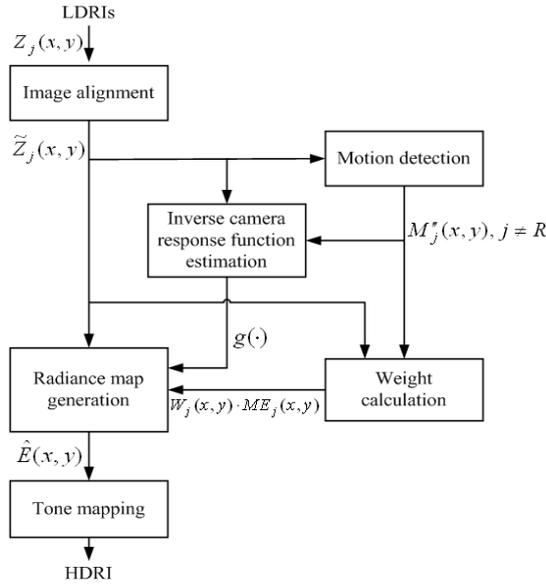
Figure 1. Block diagram of the proposed method in HDR imaging.

## 4.1. Saturated Region Separation

Saturated region separation is required for preventing false detection that can produce artifacts in a final HDRI. When false detection occurs, motion region contains the pixels that are not corrupted by a moving object. Thus, some valuable irradiance information may be lost during the process of generating the HDRI. The regions where false detection mainly occurs are over-/under-saturated in the reference image. Therefore, before finding the motion regions, we need to divide the reference image into the over-/under-saturated and non-saturated regions as

$$
\begin{aligned}
O &= \{(x, y) \mid \tilde{L}_R(x, y) \geq 254\}, \\
U &= \{(x, y) \mid \tilde{L}_R(x, y) \leq 10\} \\
N &= \{(x, y) \mid 10 < \tilde{L}_R(x, y) < 254\}, \\
O &\cap U \cap N = \phi
\end{aligned}
\tag{2}
$$

where $O$ denotes the set of pixels that gives the over-saturated regions and $U$ signifies the set of pixels that gives the under-saturated regions. $N$ represents the set of pixels that gives the non-saturated regions. Performance depends on selection of threshold values, with different values for images with different characteristics. They are experimentally selected so that the number of measurements used for radiance map generation is as large as possible with the ghost artifacts minimized. Figure 2(a) shows three regions classified in the reference image (white: over-saturated regions, gray: non-saturated regions, and black: under-saturated regions).

Because result of motion detection is unreliable in over-/under-saturated regions, we do not find motion regions in these regions. Note that the subscript $j$ in $\tilde{L}_j(x, y)$ denotes the image index, in which large $j$ represents longer exposure time. For $j \geq R$ and $(x, y) \in O$, $\tilde{L}_j(x, y)$ appears as over-saturated values, which do not contribute to extending dynamic range. However, for $j < R$ and $(x, y) \in O$, $\tilde{L}_j(x, y)$ have irradiance values (not saturated) that are accurately recorded at lower exposure setting than that of the reference image. It is possible that the images that are exposed shorter than the reference image contain a lot of details for the regions corresponding to the over-saturated regions in the reference image. We have to preserve the details in order to extend the

dynamic range of the regions. However, if we find motion in over-saturated regions, some of details are falsely detected as motion and may not be used for generating an HDRI.

To exclude unreliable measurements in generating a final radiance map, we do not find object motion in the saturated region and use all the irradiance values in $\tilde{L}_j(x,y)$, $j < R$, $(x,y) \in O$, to make the final radiance map $\hat{E}(x,y)$. Also, we ignore the irradiance values of pixels in $\tilde{L}_j(x,y)$, $j > R$, $(x,y) \in O$, in the longer exposure images because those pixels are also saturated.

## 4.2. Feature Map Generation

After separating saturated regions from the reference image $\tilde{L}_R(x,y)$, we need to find motions between the non-saturated regions in the reference LDRI and the corresponding regions in the other LDRIs. If we apply the conventional motion detection algorithms to differently exposed images, intensity shift or variation caused by illumination change makes it difficult to find accurate motions. Therefore, it is difficult to discriminate between motions caused by a moving object and motions caused by illumination changes when two differently exposed images are compared. We tackle the problem by using illumination invariant features that do not change its value under varying illuminations. Edge maps [18], image intensity derivations [19], Gabor-filter based features [20] and the zero-mean normalized cross correlation (ZNCC) [21] have been proposed as features robust to illumination changes. In the proposed motion detection method, we use the ZNCC value as a feature. It is also possible to use other features to find correspondences between LDRIs. The ZNCC has been widely used as a function to measure the degree of similarity between two image patches due to its robustness to illumination changes [21–23]. Two patches are compared after luminance and variance normalization so that the ZNCC is robust to illumination variations. The ZNCC is basically computed pixel by pixel, which means that it can capture substantial shape differences and pixel-wise variations between aligned LDRIs. Therefore, we use the ZNCC value at each pixel as a feature to decide whether the pixel is located in motion region or static region. We generate $J-1$ feature maps ($j \neq R$), which are defined in (3), where $\tilde{\mu}_R(x,y)$ and $\tilde{\mu}_j(x,y)$ are the sample averages of the windowed sub-images ($s \times s$) centered at ($x$, $y$) in the reference LDRI and $j$-th ($j \neq R$) aligned LDRIs, respectively. The square window size $s$ is assumed to be odd, e.g., selected experimentally as 11 in our experiments. Note that feature values are computed only in non-saturated regions.
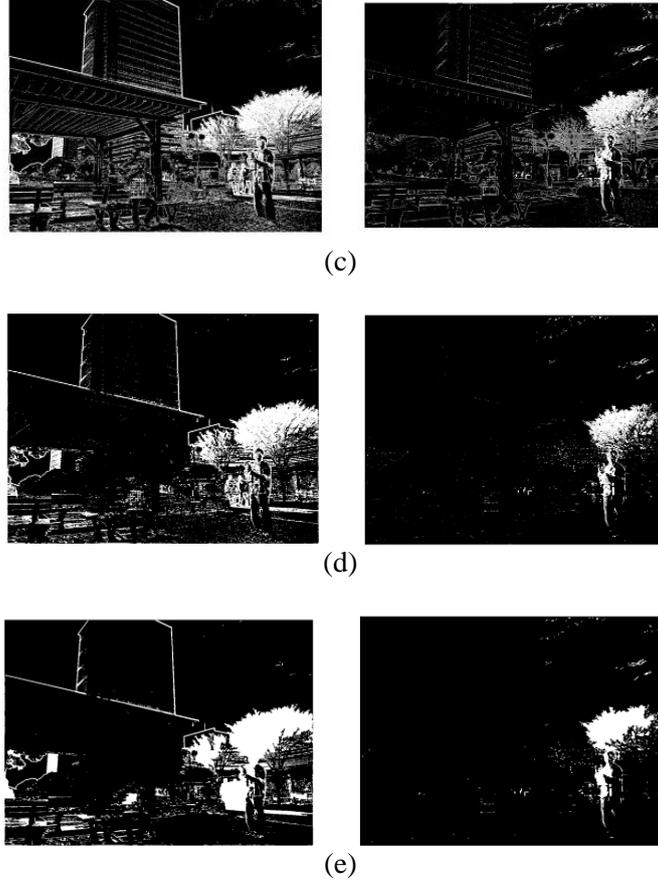


(a)



(b)

(c)



(d)



(e)
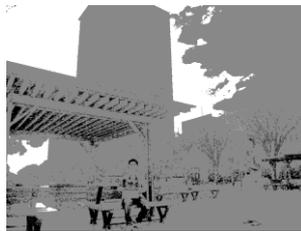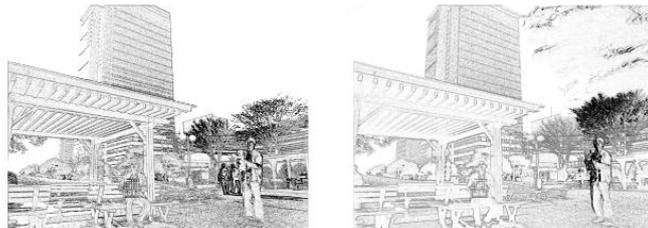
Figure 2. Intermediate results of the motion detection (Building). (a) saturated regions in $\widetilde{L}_2(x, y)$, (b) feature maps $\left|F_1(x, y)\right|$ and $\left|F_3(x, y)\right|$ generated from the luminance images, (c) motion maps $M_1(x, y)$ and $M_3(x, y)$ by thresholding of (b), (d) motion maps $M_1'(x, y)$ and $M_3'(x, y)$ by reclassifying undefined pixels in (c), (e) final motion maps $M_1''(x, y)$ and $M_3''(x, y)$ after motion map clustering.

$$F_j(x, y) = \begin{cases} \dfrac{\displaystyle\sum_{u=-(s-1)/2}^{(s-1)/2}\sum_{v=-(s-1)/2}^{(s-1)/2}\left[\widetilde{L}_R(x+u, y+v) - \widetilde{\mu}_R(x, y)\right]\cdot\left[\widetilde{L}_j(x+u, y+v) - \widetilde{\mu}_j(x, y)\right]}{\sqrt{\displaystyle\sum_{u=-(s-1)/2}^{(s-1)/2}\sum_{v=-(s-1)/2}^{(s-1)/2}\left[\widetilde{L}_R(x+u, y+v) - \widetilde{\mu}_R(x, y)\right]^2} \cdot \sqrt{\displaystyle\sum_{u=-(s-1)/2}^{(s-1)/2}\sum_{v=-(s-1)/2}^{(s-1)/2}\left[\widetilde{L}_j(x+u, y+v) - \widetilde{\mu}_j(x, y)\right]^2}}, & (x, y) \in N \\[2em] 1, & \text{otherwise} \end{cases}$$

(3)

Figure 2(b) shows the feature maps using the luminance images. The ZNCC value at each pixel ranges from −1 to 1. If a pixel is located in a moving object, the ZNCC value will be close to 0. On the contrary, if a pixel is located in a static region, it will be close to 1 or −1. Although most of the pixels in motion region have low ZNCC values, some pixels in motion region have high ZNCC values. This situation occurs, especially when the pixels are located in homogeneous part of the object that moves across homogenous part of the static region. In homogenous regions, there is not enough structural information to detect motion between given patches. Thus, moving

pixels may be misclassified into the non-moving pixels. We solve this problem by adaptive local thresholding that gives more accurate classification results.

## 4.3. Adaptive Local Thresholding

Feature maps $\left|F_j(x, y)\right|$ $(j \neq R)$ need to be converted into motion maps $M'_j(x, y)$ $(j \neq R)$ that classify a motion/static region by thresholding with a pre-defined global threshold $T_G$. By choosing an appropriate global threshold value, we can divide the entire image into two regions: motion region and static region. The disadvantage of global thresholding is that a single threshold value determines the accuracy of a motion map $M'_j(x, y)$ $(j \neq R)$. There are two cases when false detection occurs. First, non-moving pixels are incorrectly detected as moving pixels (e.g., edges of the building in Figure 2(b)) when $T_G$ is too large. Second, moving pixels are incorrectly detected as non-moving pixels (e.g., backs of two persons in the background in Figure 2(b)) when $T_G$ is too small. Therefore, global thresholding with a single threshold has a limitation on detecting the motion region accurately. For better motion detection performance, we present a local thresholding method that adapts the threshold value at each pixel according to the local image characteristics.

Canny used thresholding with hysteresis for edge detection [24]. Chen *et al*. proposed a two-stage approach of entropic thresholding based on 2-D histogram, providing a good quality of segmentation [25]. Yan *et al*. proposed a multistage adaptive thresholding method for image segmentation [26]. They first segmented an image by global thresholding and made use of local image statistics of mean and variance in neighborhood to obtain local threshold for each pixel.

Adaptive local thresholding proposed in this paper is described as follows. After generating $J$–1 feature maps $\left|F_j(x, y)\right|$ $(j \neq R)$, we classify the whole pixels into three groups: 1) moving pixels, 2) undefined pixels, and 3) non-moving pixels. We begin by introducing two thresholds: low threshold $T_L$ and high threshold $T_H$. Figure 2(c) shows tri-level motion maps $M_j(x, y)$ $(j \neq R)$ of thresholding by $T_L$ and $T_H$. Pixels whose feature value is less than low threshold $T_L$ are classified as moving pixels (white) whereas pixels whose feature value is larger than high threshold $T_H$ are classified as non-moving pixels (black). Pixels whose feature value is larger than $T_L$ and smaller than $T_H$ are classified as undefined pixels (gray). The use of two thresholds with hysteresis gives more flexibility than a single global threshold approach.

However, there is still a problem related to selection of optimal threshold values of $T_L$ and $T_H$ that work well on various images of different characteristics. We observe that the pixels having feature values $\left|F_j(x, y)\right|$ $(j \neq R)$ between 0.6 and 0.8 have a high probability of being misclassified. Therefore, we set $T_L$ to 0.6 to assure that the pixels having whose feature values are smaller than 0.6 are moving pixels. In the same manner, we set $T_H$ to 0.8 to assure that the pixels whose feature values are larger than 0.8 are non-moving pixels. The undefined pixels, whose feature values are between 0.6 and 0.8, need to be reclassified into moving and non-moving pixels.
We notice that moving pixels are clustered by a moving object. Undefined pixels located in motion regions where moving pixels are densely populated (e.g., two persons in the background and branches of trees in Figure 2(c)), are likely to have feature values larger than those of neighboring pixels, which are smaller than $T_L$. Whereas undefined pixels located in static regions where non-moving pixels and undefined pixels are a majority tend to have smaller feature values than those of neighboring pixels, which are larger than $T_L$. Thus, if we adaptively select threshold value $T_j(x, y)$ $(j \neq R)$ by considering ZNCC feature values of neighboring pixels, we can

correctly reclassify undefined pixels into moving pixels and non-moving pixels. Based on this observation, the threshold value for undefined pixels is set to the local mean. We define the pixelwise threshold value for the feature value in the non-saturated region of each LDRI ($j \neq R$) as

$$T_j(x, y) = \begin{cases} T_L - |F_j(x, y)|, & \text{for } |F_j(x, y)| \leq T_L \\ T_H - |F_j(x, y)|, & \text{for } |F_j(x, y)| \geq T_H \\ ||F_j(x, y)| - \mu_j(x, y), & \text{for } T_L < |F_j(x, y)| < T_H \end{cases} \quad , (x, y) \in N \quad (4)$$

where $j \neq R$ and $\mu_j(x, y)$ is the mean of absolute ZNCC values of neighboring pixels in $ws \times ws$ square window centered at $(x, y)$. Window size $ws$ is a user-defined parameter, which affects the accuracy of local thresholding. A too small window size makes local thresholding sensitive to noise and false detection. On the other hand, a too large window size increases the computational complexity. We change window size $ws$ and select the window size experimentally as $13 \times 13$ for reliable local thresholding.

After calculating $T_j(x, y)$ ($j \neq R$) using (4) for all the pixels in non-saturated regions, we generate $J-1$ binary motion maps $M'_j(x, y)$ ($j \neq R$) by

$$M'_j(x, y) = \begin{cases} 1, & \text{for } (x, y) \in N \text{ and } T_j(x, y) \geq 0 \\ 0, & \text{for } (x, y) \in N \text{ and } T_j(x, y) < 0 \\ 0, & \text{for } (x, y) \in O \text{ and } j < R \\ 1, & \text{for } (x, y) \in O \text{ and } j > R \\ 1, & \text{for } (x, y) \in U \text{ and } j < R \\ 0, & \text{for } (x, y) \in U \text{ and } j > R \end{cases} \quad (5)$$

where pixels labeled as 1 correspond to those in motion regions while pixels labeled as 0 correspond to those in static regions. Based on the threshold value $T_j(x, y)$ ($j \neq R$), we determine whether a pixel is in motion regions or not. In order to exclude unreliable measurements, we deal with the over-/under-saturated regions in different ways as follows. We ignore the irradiance values of pixels in $\tilde{L}_j(x, y)$, $j > R$, $(x, y) \in O$, in the longer exposure images by labeling the corresponding pixels $M'_j(x, y)$ ($j \neq R$) by 1. Whereas we use all the irradiance values in $\tilde{L}_j(x, y)$, $j < R$, $(x, y) \in O$, in the shorter exposure images by labeling the corresponding pixels $M'_j(x, y)$ ($j \neq R$) by 0. Similarly, the under-saturated regions are handled. Figure 2(d) shows the result of motion maps $M'_1(x, y)$ and $M'_3(x, y)$ where we can see that false detection is decreased.

## 4.4. Motion Map Clustering

Although the moving pixels are well clustered, there still remain some isolated pixels and holes inside the moving object in the motion maps $M'_j(x, y)$, ($j \neq R$). In order to remove ghost artifacts clearly, we need to eliminate the isolated pixels and fill the holes in the moving object. First, we remove isolated pixels that are surrounded by pixels labeled as 0. Then, we link the pixels that are separated by one-pixel gaps and fill the holes in the moving object by using a fast hybrid reconstruction algorithm proposed in [27]. We extract the connected components of the motion map and perform iteratively dilation from boundaries of the image to the connected components

until further dilation ceases to change the motion map. We use the motion map as a mask image. Marker images are labeled as 0 except that the boundary pixels (starting points) are labeled as 1. Figure 2(e) shows the results of the final motion maps $M_1''(x, y)$ and $M_3''(x, y)$ after applying the morphological reconstruction algorithm. We can see that isolated false detection is decreased and the pixels in holes are clustered to adjacent motion regions.

## 4.5. Radiance Map Generation

Once the final motion map $M_j''(x, y)$ ($j \neq R$) is found, we can reduce ghost artifacts. HDRI generation proceeds depending on whether the pixel is included in the motion region or not. For the former case, the weight for the pixels that are inside the motion region is set to zero so that the irradiance values at these pixels do not contribute to computation of the final radiance value. For the latter case, the conventional method [2] is used. The final radiance map is a weighted average of the corresponding irradiance values obtained from aligned LDRIs in the logarithm domain for each color channel independently. Based on Debevec and Malik's method [2], we describe the ghost-free HDRI generation process as follows:

$$\ln \hat{E}(x, y) = \frac{\sum_{j=1}^{J} \left( W_j(x, y) \right)\left( ME_j(x, y) \right)\left( \ln g(\tilde{Z}_j(x, y)) - \ln \Delta t_j \right)}{\sum_{j=1}^{J} \left( W_j(x, y) \right)\left( ME_j(x, y) \right)} \qquad (6)$$

where $g(\cdot)$ is the inverse camera response function and $\Delta t_j$ is the exposure time of $j$-th aligned LDRI. $J$ denotes the total number of LDRIs and $\tilde{Z}_j(x, y)$ represents the intensity of $j$-th aligned LDRI. The weighting factor is defined as

$$ME_j(x, y) = \begin{cases} 1, & \text{for } j = R \\ 0, & \text{for } M_j''(x, y) = 1 \text{ and } j \neq R \\ 1, & \text{for } M_j''(x, y) = 0 \text{ and } j \neq R \end{cases} \qquad (7)$$

where $M_j''(x, y)$ ($j \neq R$) is the final motion map. We generate motion maps in luminance channel and the weighting factor is used for each color channel. Several weighting functions have been proposed to produce better results by increasing the weights of the reliable pixels while decreasing those of the noisy and saturated pixels [2, 28, 29]. Debevec and Malik proposed a hat-shaped weighting function where the pixels in the middle of the intensity range are given higher weights and those near extremes get lower weights. They assumed that the pixels in the middle range have more reliable irradiance values. We generate $J$ weight maps using a hat-shaped weighting function proposed by Debevec and Malik [2]. With image alignment, the modified weight map is defined as

$$W_j(x, y) = \begin{cases} \tilde{Z}_j(x, y) - \tilde{Z}_j^{\min}, & \text{for } \tilde{Z}_j(x, y) \leq \frac{1}{2}\left( \tilde{Z}_j^{\min} + \tilde{Z}_j^{\max} \right) \\ \tilde{Z}_j^{\max} - \tilde{Z}_j(x, y), & \text{otherwise} \end{cases} \qquad (8)$$

where $\tilde{Z}_j(x, y)$ is the intensity of $j$-th aligned LDRIs. $\tilde{Z}_j^{\min}$ and $\tilde{Z}_j^{\max}$ denote the minimum and maximum intensity values of $\tilde{Z}_j(x, y)$ for each color channel, respectively.

## 5. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, we discuss how LDRI capture setting can influence the final HDRI result in terms of ghost removal. Then, we provide experimental results of the proposed ghost removal method and compare them with those of existing methods. All HDRIs are created using a number of LDRIs captured with hand-held cameras with different exposure time ranging from 1/8000 s to 1/50 s.

### 5.1. Image Capture Setting and Discussions

LDRIs can be captured with various settings: exposure time, number of images, and EV spacing. Generally, more LDRIs with small EV spacing guarantee that there are less artifacts and noise in an HDRI [6, 30]. One of the advantages generating an HDRI with many LDRIs is that we can obtain multiple irradiance values or measurements at each pixel between LDRIs so that the SNR of the final radiance map $\hat{E}(x, y)$ can be improved [6, 31]. However, a large exposure set accompanies the increased processing time, memory requirement, and the increased acquisition time. These are major restrictions on consumer applications such as a digital camera and mobile phone which have a limited computing power. Moreover, as the total capture time increases, object motion would be larger and more frequent so that more ghost artifacts will appear in an HDRI. Therefore, we mainly generate HDRIs using LDRIs with three different exposure times ($J$=3) to apply consumer applications.

In general, LDRIs with 2 EV spacing can capture a wider dynamic range than LDRIs with 1 EV spacing. However, a large EV spacing is necessarily undesirable because large illumination changes between LDRIs can decrease the accuracy of motion detection. Let us consider the case in which a static region is detected as motion region because of large intensity changes due to large illumination changes. In this case, the information of irradiance at pixels in the false detection region is not used for the merging process. This can result in the dynamic range decrease and the SNR degradation in the final HDRI. However, a large EV spacing is inevitable to obtain a wider dynamic range with the fixed number of LDRIs, thus it is important to minimize the possibility of false detection in LDRIs with a large EV spacing.

### 5.2. Performance Comparison of Ghost Removal Methods

We compare the proposed method with three existing methods: Jacobs *et al.*'s method [13], Khan *et al.*'s method (10 iterations) [9], and Min *et al.*'s method [14]. For a fair comparison, a SIFT based alignment method [14] is used to generate aligned LDRIs as input of the proposed method and three existing methods. In addition, the camera response estimation method by Debevec and Malik [2] and tone-mapping method by Reinhard *et al.* [32] are used in all experiments. In experiments, we select a mid-exposure image as the reference image, because, in general, over-/under-saturated region is the least in the mid-exposure image. Although three exposure images are used in experiments, these methods including the proposed method can cope with more exposures such as 5, 7, or 9. This paper shows results of the proposed and existing methods in general case (Figure 3) and more complex cases (Figures 4–6).

Figures 3(a)–(e) show the HDRI by conventional method [2] without ghost removal, and ghost artifacts reduced HDRIs by three existing methods [13, 9, 14] and proposed method. All HDRIs are generated from the aligned version of LDRIs. For each image, left image shows the HDRI, in which the regions of ghost artifacts are marked with boxes. Cropped images of these regions are shown in top right and bottom right. As shown in Figures 3(b) and 3(c), Jacobs *et al.*'s method and Khan *et al.*'s method do not remove two persons walking in the background and the artifacts

on the face of a man standing in front of the building. In Figure 3(d), Min *et al.*'s method removes most of the ghost artifacts, while some artifacts caused by false detection or saturated regions (clouds) still remain. Figure 3(e) shows that the proposed method gives better results than three existing methods in terms of ghost removal. The proposed method clearly removes ghost artifacts caused by small motion of a man and occlusion.



Figure 3. Performance comparison of ghost removal methods (Building, left: HDRI, right: cropped images). (a) conventional method [2] (without ghost removal), (b) Jacobs *et al.*'s method [13], (c) Khan *et al.*'s method [9], (d) Min *et al.*'s method [14], (e) proposed method.

Figure 4(a) shows a set of aligned LDRIs (Cafeteria, size: 2192×1644, *J*=3, *R*=2, exposure time (from left to right): 1/1280 s, 1/320 s, 1/80 s). Figures 4(b)–4(f) show the HDRI by conventional method [2] without ghost removal, and ghost artifacts reduced HDRIs by three existing methods [13, 9, 14] and proposed method. In Figures 4(b)–4(f), left image shows the HDRIs, in which the regions of ghost artifacts are marked with boxes. Cropped images of those regions are shown on the top right and bottom right. The ghost artifacts are shown on windows that reflect the crowd with various motions. This is a special case because the motions include reflections and low contrast regions altogether. Note that the proposed method clearly removes the ghost artifacts, as shown in Figure 4(f). In three existing methods, there still remain ghost artifacts as well as artifacts on the texture of the floor caused by false detection in saturated regions of the reference image.
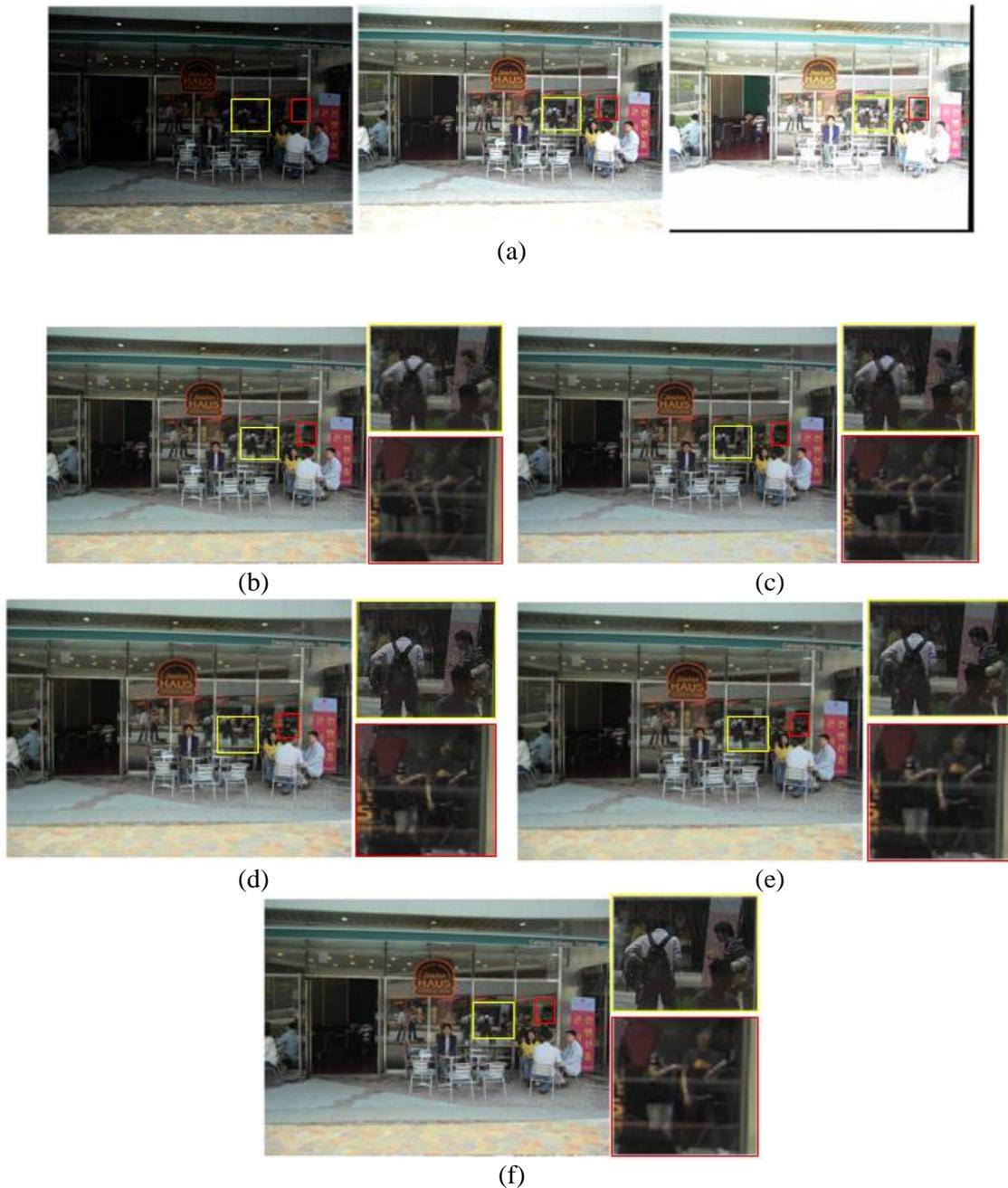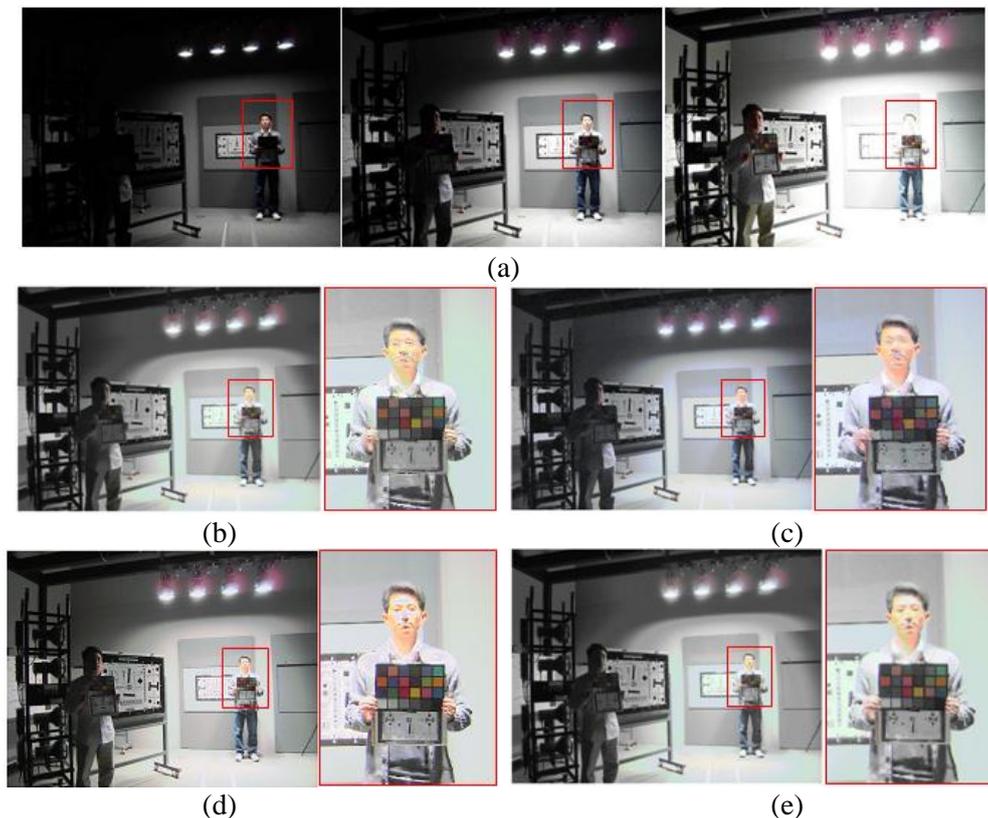
(a)



(b)                                                      (c)



(d)                                                      (e)



(f)

Figure 4. Performance comparison of ghost removal methods. (a) a set of aligned LDRIs $\widetilde{Z}_1(x, y), \widetilde{Z}_2(x, y),$ and $\widetilde{Z}_3(x, y)$ (Cafeteria, size: 2192×1644, $J = 3$, $R = 2$, exposure time (from left to right): 1/1280 s, 1/320 s, 1/80 s), in (b)–(f), left: HDRI, right: cropped images from the HDRI, (b) conventional method [2] (without ghost removal), (c) Jacobs *et al.*'s method [13], (d) Khan *et al.*'s method [9], (e) Min *et al.*'s method [14], (f) proposed method.

Figure 5(a) shows a set of aligned LDRIs of an indoor scene (Laboratory, size: 2192×1644, $J$=3, $R$=2, exposure time (from left to right): 1/2000 s, 1/500 s, 1/125 s). Figures 5(b)–5(f) show the HDRI by conventional method [2] without ghost removal, and ghost artifacts reduced HDRIs by three existing methods [13, 9, 14] and proposed method. In Figures 5(b)–5(f), left images show the HDRIs. Right images show cropped images of those regions. In this set of aligned LDRIs, a man in the right side of the image moves and the lamps that are strongly illuminated are located right over a man's head. False detection is likely to occur when the moving object (a man) is

influenced by strong illumination changes. In the three existing methods, false contour is visible inside the face of a man. In addition, some artifacts still remain in a man's hands as shown in Figures 5(b) and 5(c). The proposed method clearly removes both artifacts. This is because we ignore the irradiance values of pixels in saturated regions and do not find motions in a man's face. Figure 5(a) shows a good example. There are lens-flare artifacts around an illuminator in $\tilde{Z}_2(x, y)$ and $\tilde{Z}_3(x, y)$. Since we select $\tilde{Z}_2(x, y)$ as a reference image to generate an HDRI, lens-flare artifacts are still shown in the HDRIs in Figures 5(b)–5(f). In $\tilde{Z}_1(x, y)$, there are not lens-flare artifacts around an illuminator. Thus, if we remove the artifacts in the HDRI, this part of $\tilde{Z}_1(x, y)$ can be used as the reference instead of the region in the reference image $\tilde{Z}_2(x, y)$.

Figure 6(a) shows a set of aligned LDRIs with large motion (Warehouse, size: 2336×1552, $J$=3, $R$=2, exposure time (from left to right): 1/4000 s, 1/1500 s, 1/350 s). The LDRIs are affected by camera rotation as well as translation. We align $Z_1(x, y)$ and $Z_3(x, y)$ with respect to the reference LDRI $Z_2(x, y)$, using the SIFT based method [16], which this alignment causes the black regions in boundaries in $\tilde{Z}_1(x, y)$ and $\tilde{Z}_3(x, y)$, and the green regions in the boundaries of Figures 6(b)–(f). Figures 6(b)–6(f) show the HDRI by conventional method [2] without ghost removal, and ghost artifacts reduced HDRIs by three existing methods [13, 9, 14] and proposed
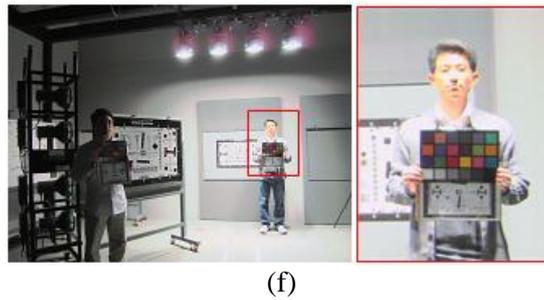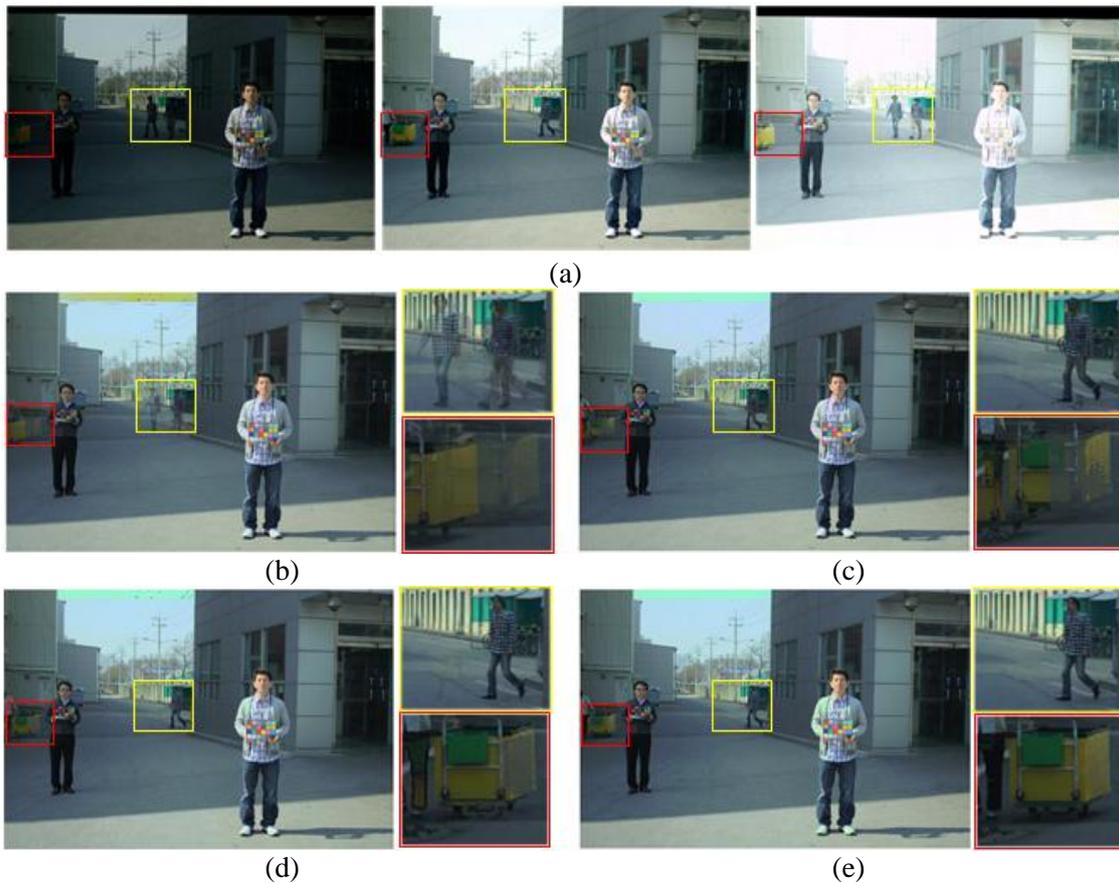

(a)


(b)


(c)


(d)


(e)

(f)

Figure 5. Performance comparison of ghost removal methods. (a) a set of aligned LDRIs $\widetilde{Z}_1(x, y), \widetilde{Z}_2(x, y),$ and $\widetilde{Z}_3(x, y)$ (Laboratory, size: 2192×1644, $J = 3$, $R = 2$, exposure time (from left to right): 1/2000 s, 1/500 s, 1/125 s), in (b)–(f), left: HDRI, right: a cropped image from the HDRI, (b) conventional method [2] (without ghost removal), (c) Jacobs *et al.*'s method [13], (d) Khan *et al.*'s method [9], (e) Min *et al.*'s method [14], (f) proposed method.

method. In Figures 6(b)–6(f), left images show the HDRIs, in which the regions of ghost artifacts are marked with two boxes (two persons walking in the background and a person standing next to a garbage wagon). Right images show cropped images of those regions. Ghost artifacts shown in Figure 6(b) are removed in Khan *et al.*'s method and Min *et al.*'s method as well. However, ghost artifacts still remain inside a person and garbage wagon as shown in Figures 6(c)–6(e). The proposed method clearly removes both artifacts as shown in Figure 6(f).
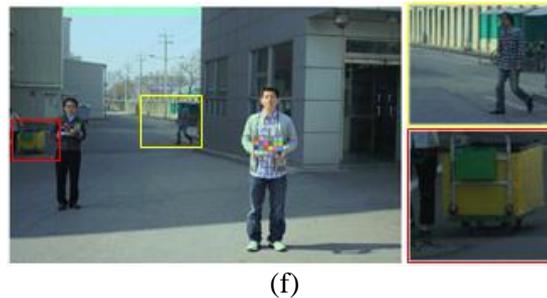


(a)



(b)



(c)



(d)



(e)

(f)

Figure 6. Performance comparison of ghost removal methods. (a) a set of aligned LDRIs $\tilde{Z}_1(x, y), \tilde{Z}_2(x, y),$ and $\tilde{Z}_3(x, y)$ (Warehouse, size: 2336×1552, $J$ = 3, $R$ = 2, exposure time (from left to right): 1/4000 s, 1/1500 s, 1/350 s), in (b)–(f), left: HDRI, right: cropped images from the HDRI, (b) conventional method [2] (without ghost removal), (c) Jacobs *et al.*'s method [13], (d) Khan *et al.*'s method [9], (e) Min *et al.*'s method [14], (f) proposed method.

## 6. CONCLUSIONS

We propose a feature based ghost removal method in HDRI generation. In order to minimize the influence of saturated regions in ghost removal process, we detect non-saturated regions in the reference LDRI. For the non-saturated regions, we find correspondences between LDRIs using the ZNCC and then find object motion region through post processing such as adaptive local thresholding and motion clustering. We also concentrate on reducing false detection regions because false detection decreases the dynamic range of the HDRI and causes artifacts in the HDRI. We discuss the effect of EV setting as well as the number of LDRI for HDRI generation. Experiments with LDRIs sets captured with different exposure time show that the proposed method gives better performance than the existing methods. Future work will focus on the selection of the self-consistent reference image.

## ACKNOWLEDGEMENTS

## REFERENCES

[1]   M. D. Tocci, C. Kiser, N. Tocco, and P. Sen, "A versatile HDR video production system," ACM Trans. Graphics, vol. 30, no. 4, article no. 41, Jul. 2011.

[2]   P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in Proc. SIGGRAPH 97, pp. 369–378, Los Angeles, CA, Aug. 1997.

[3]   B.-D. Choi, S.-W. Jung, and S.-J. Ko, "Motion-blur-free camera system splitting exposure time," IEEE Trans. Consumer Electronics, vol. 54, no. 3, pp. 981−986, Aug. 2008.

[4]   W. Zhang and W.-K. Cham, "Gradient-directed composition of multi-exposure images," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 530−536, San Francisco, CA, Jun. 2010.

[5]   E. Reinhard, G. Ward, P. Debevec, and S. Pattanaik, High Dynamic Range Imaging: Acquisition, Display, and Image Based Lighting. 2nd ed., San Francisco, CA: Morgan Kaufmann, 2010.

[6]   S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High dynamic range video," ACM Trans. Graphics, vol. 22, no. 3, pp. 319–325, Jul. 2003.

[7]   G. Ward, "Fast, robust image registration for compositing high dynamic range photographs for handheld exposures," Journal of Graphics Tools, vol. 8, no. 2, pp. 17−30, Jan. 2003.

[8] N. Barakat, A. N. Hone, and T. E. Darcie, "Minimal-bracketing sets for high dynamic range image capture," IEEE Trans. Image Processing, vol. 17, no. 10, pp. 1864–1875, Oct. 2008.

[9] E. A. Khan, A. O. Akyuz, and E. Reinhard, "Robust generation of high dynamic range images," in Proc. Int. Conf. Image Processing, pp. 2005–2008, Atlanta, GA, Oct. 2006.

[10] M. Pedone and J. Heikkila, "Constrain propagation for ghost removal in high dynamic range images," in Proc. Third Int. Conf. Computer Vision Theory and Applications, pp. 36–41, Madeira, Portugal, Jan. 2008.

[11] O. Gallo, N. Gelfand, W. Chen, and M. Tico, "Artifact-free high dynamic range imaging," in Proc. IEEE Int. Conf. Computational Photography, San Francisco, CA, Apr. 2009.

[12] T. Jinno and M. Okuda, "Motion blur free HDR image acquisition using multiple exposures," in Proc. IEEE Int. Conf. Image Processing, pp. 1304–1307, San Diego, CA, Oct. 2008.

[13] K. Jacobs, C. Loscos, and G. Ward, "Automatic high dynamic range image generation for dynamic scenes," IEEE Computer Graphics and Applications, vol. 28, no. 2, pp. 84–93, Mar. 2008.

[14] T.-H. Min, R.-H. Park, and S. Chang, "Histogram based ghost removal in high dynamic range images," in Proc. IEEE Int. Conf. Multimedia and Expo, pp. 530−533, New York, Jun./Jul. 2009.

[15] J. Im, J. Jeon, M. Hayes, and J. Paik, "Single image-based ghost-free high dynamic range imaging using local histogram stretching and spatially-adaptive denoising," IEEE Trans. Consumer Electronics, vol. 57, no. 4, pp. 1478−1484, Nov. 2011.

[16] M. D. Grossberg and S. K. Nayar, "What is the space of camera response functions?," in Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition, pp. 602–609, Madison, WI, Jun. 2003.

[17] A. Tomaszewska and R. Mantiuk, "Image registration for multi-exposure high dynamic range image acquisition," in Proc. Int. Conf. Central Europe Computer Graphics, Visualization and Computer Vision, pp. 49–56, Plzen-Bory, Czech, Jul. 2007.

[18] X. Yong, S. Gao, and M. K. H. Leung, "Face recognition using line edge map," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 6, pp. 764–779, Jun. 2002.

[19] A. Shashua and T. Riklin, "The quotient image: Class-based re-rendering and recognition with varying illuminations," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 2, pp. 129–139, Feb. 2001.

[20] J. K. Kamarainen, V. Kyrki, and H. Kalviainen, "Invariance properties of Gabor filter based features– overview and applications," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 28, no. 5, pp. 1088–1099, May 2006.

[21] F. Tombari and L. D. Stefano, "A robust measure for visual correspondence," in Proc. IEEE Int. Conf. Image Analysis and Processing, pp. 376–381, Modena, Italy, Sep. 2007.

[22] D. M. Tsai, C. T. Lin, and J. F. Chen, "The evaluation of normalized cross correlations for defect detection," Pattern Recognition Letters, vol. 24, no. 15, pp. 2525–2535, Nov. 2003.

[23] L. D. Stefano, S. Mattoccia, and F. Tombari, "ZNCC-based template matching using bounded partial correlation," Pattern Recognition Letters, vol. 26, no. 14, pp. 2129–2134, Oct. 2005.

[24] J. Canny, "A computational approach to edge detection," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 8, no. 6, pp. 679–714, Nov. 1986.

[25] W. T. Chen, C. H. Wen, and C. W. Yang, "A fast two-dimensional entropic thresholding algorithm," Pattern Recognition, vol. 27, no. 7, pp. 885–893, Jul. 1994.

[26] F. Yan, H. Zhang, and C. R. Kube, "A multistage adaptive thresholding method," Pattern Recognition Letters, vol. 26, no. 8, pp. 1183–1191, Jun. 2005.

[27] L. Vincent, "Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms," IEEE Trans. Image Processing, vol. 2, no. 2, pp. 176–201, Apr. 1993.

[28] A. O. Akyuz and E. Reinhard, "Noise reduction in high dynamic range imaging," Journal of Visual Communication and Image Representation, vol. 18, no. 5, pp. 366–376, Oct. 2007.

[29] A. A. Bell, C. Seiler, J. N. Kaftan, and T. Aach, "Noise in high dynamic range imaging," in Proc. Int. Conf. Image Processing, pp. 561–564, San Diego, CA, Oct. 2008.

[30] M. D. Grossberg and S. K. Nayar, "High dynamic range from multiple images: Which exposures to combine?," in Proc. ICCV Workshop Color and Photometric Methods in Computer Vision, Nice, France, Oct. 2003.

[31] N. Barakat, T. E. Darcie, and A. N. Hone, "The tradeoff between SNR and exposure-set size in HDR imaging," in Proc. IEEE Int. Conf. Image Processing, pp. 1848–1851, San Diego, CA, Oct. 2008.

[32] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," ACM Trans. Graphics, vol. 21, no. 3, pp. 267–276, Jul. 2002.

## Authors

Hwan-Soon Sung received the B.S. and M.S. degrees in electronic engineering from Sogang University in 2008 and 2010, respectively. His current research interests are image processing and image enhancement.

Rae-Hong Park received the B.S. and M.S. degrees in electronics engineering from Seoul National University, Seoul, Korea, in 1976 and 1979, respectively, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1981 and 1984, respectively. In 1984, he joined the faculty of the Department of Electronic Engineering, School of Engineering, Sogang University, Seoul, Korea, where he is currently a Professor. In 1990, he spent his sabbatical year as a Visiting Associate Professor with the Computer Vision Laboratory, Center for Automation Research, University of Maryland at College Park. In 2001 and 2004, he spent sabbatical semesters at Digital Media Research and Development Center (DTV image/video enhancement), Samsung Electronics Co., Ltd., Suwon, Korea. In 2012, he spent a sabbatical year in Digital Imaging Business (R&D Team) and Visual Display Business (R&D Office), Samsung Electronics Co., Ltd., Suwon, Korea. His current research interests are video communication, computer vision, and pattern recognition. He served as Editor for the Korea Institute of Telematics and Electronics (KITE) *Journal of Electronics Engineering* from 1995 to 1996. Dr. Park was the recipient of a 1990 Post-Doctoral Fellowship presented by the Korea Science and Engineering Foundation (KOSEF), the 1987 Academic Award presented by the KITE, the 2000 Haedong Paper Award presented by the Institute of Electronics Engineers of Korea (IEEK), the 1997 First Sogang Academic Award, and the 1999 Professor Achievement Excellence Award presented by Sogang University. He is a co-recipient of the Best Student Paper Award of the IEEE Int. Symp. Multimedia (ISM 2006) and IEEE Int. Symp. Consumer Electronics (ISCE 2011).

Dong-Kyu Lee received the B.S. and M.S. degrees in electronic engineering from Sogang University in 2010 and 2012, respectively. His current research interests are image processing and image enhancement.

SoonKeun Chang received his B.S. degree in astronomy and space science from KyungHee University, Korea, in 2000 and M.S. degree in control engineering from Kanazawa University in 2003. He received a Ph.D. degree in control engineering from Tokyo Institute of Technology (TITech) in 2007. Now he works at Samsung Electronics Co., Ltd., Korea. His main research interests include computer vision and image processing.