EUROGRAPHICS 2016 / J. Jorge and M. Lin
(Guest Editors)

Volume 35 (2016), Number 2

An Objective Deghosting Quality Metric for HDR Images

Okan Tarhan Tursun[1], Ahmet Oğuz Akyüz[1], Aykut Erdem[2] and Erkut Erdem[2]

[1]Dept. of Computer Engineering, Middle East Technical University, Turkey
[2]Dept. of Computer Engineering, Hacettepe University, Turkey

(a) Moving people generate blending (red) and visual difference (blue) artifacts. (b) Over-smoothing gives rise to gradient inconsistency (green) artifacts.

Figure 1: Our metric detects several kinds of HDR deghosting artifacts. In (a), Khan et al.'s [KAR06] output is shown in the bottom-left corner and our metric's result in the bottom-right. The same for (b), except Hu et al.'s [HGPS13] deghosting algorithm is used. Exposure sequences are shown on the top. Cyan color occurs due to both gradient and visual difference metrics producing high output.

Abstract

Reconstructing high dynamic range (HDR) images of a complex scene involving moving objects and dynamic backgrounds is prone to artifacts. A large number of methods have been proposed that attempt to alleviate these artifacts, known as HDR deghosting algorithms. Currently, the quality of these algorithms are judged by subjective evaluations, which are tedious to conduct and get quickly outdated as new algorithms are proposed on a rapid basis. In this paper, we propose an objective metric which aims to simplify this process. Our metric takes a stack of input exposures and the deghosting result and produces a set of artifact maps for different types of artifacts. These artifact maps can be combined to yield a single quality score. We performed a subjective experiment involving 52 *subjects and* 16 *different scenes to validate the agreement of our quality scores with subjective judgements and observed a concordance of almost 80%. Our metric also enables a novel application that we call as hybrid deghosting, in which the output of different deghosting algorithms are combined to obtain a superior deghosting result.*

Categories and Subject Descriptors (according to ACM CCS): I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Motion

1. Introduction

Due to its low-cost and availability, the most commonly used HDR image capture method remains to be the multiple exposures technique (MET), which involves combining a set of exposures of a scene into a single HDR image [DM97]. The main requirements of this technique are that the camera and the captured scene remain static throughout the capture process. Otherwise, the lack of correspondence between exposures result in what is known as *ghosting artifacts*. While stabilizing a camera can be achieved by using a tripod, ensuring a static scene is much more difficult as most real-world scenes contain dynamic objects. Many deghosting algorithms have been proposed to address this problem ranging from

simple alignment methods to sophisticated computer vision algorithms. To this date, more than 50 deghosting algorithms have been proposed [TAEE15].

As in all fields, the proliferation of these algorithms gave rise to subjective experiments that aim to evaluate their performance [HT13, HTM13, HTM14, TAEE15]. However, subjective comparisons of HDR deghosting algorithms is problematic for several reasons. First, ideally the comparison medium must be an HDR display [SHS*04], as otherwise some artifacts may be lost or new artifacts may be generated during tone mapping. Secondly, the comparison task is challenging as participants need to compare a stack of LDR images with one or more deghosted images. Finally, the findings of subjective experiments become outdated as new algorithms are being proposed on a rapid basis.

In order to overcome these problems, there is a clear need to define objective metrics to compare HDR deghosting algorithms, which is the primary goal of this paper. A number of quality assessment metrics have been proposed for HDR images [RWPD10, Chapter 10] [SKMS06, KJF07, RFWB07, AMMS08, MKRH11]. However, none of these metrics are suitable for evaluating deghosting artifacts. The objective metric proposed in this work is the result of analyzing the outputs of several HDR deghosting algorithms to identify the most prevalent artifacts that are present. Some of these artifacts and the corresponding distortion maps produced by our algorithm are shown in Figure 1.

To validate the compliance of our metric with subjective judgements of real observers, we conducted a subjective experiment that involves 16 scenes of varying characteristics, 10 deghosting algorithms, and 52 participants. We found that there is a high degree of correlation between the subjective and objective results.

The proposed metric has several applications such as automatic comparison of deghosting algorithms, automatic image quality inspection, understanding the strengths and weaknesses of existing algorithms, optimizing parameter selection, providing feedback for developing better HDR deghosting algorithms, and hybrid deghosting in which multiple deghosting results are combined to obtain a superior one.

2. Related Work

HDR Deghosting. With more than 50 deghosting algorithms have been proposed within a decade, the development of deghosting algorithms has been one of the most active research areas within HDR imaging. HDR deghosting algorithms may be classified into three main groups as *moving object removal*, *moving object selection*, and *moving object registration* methods, in terms of the strategies used for constructing the ghost-free HDR output [TAEE15].

Moving object removal methods aim to exclude dynamic objects from the output HDR image producing the static background [KAR06, ZC10]. Moving object selection algorithms, on the other hand, first detect motion regions by inspecting the inconsistencies in the input pixel intensities and then seek to eliminate the corresponding artifacts by using samples from either a single input LDR image [JLW08, LC09, RWPD10] or a subset of the input images that are found to be consistent [GGC*09, RKC09, RC11, HLL*11, SSM12, OLK13, SPLC13, GKTT13]. Finally, moving object registration methods use local registration techniques to identify and transfer information from the best matching regions across different exposures. While some studies make use of pixel-wise optical flow correspondences [KUWS03, ST04, JO12, HDW14], others utilize patch-based dense matching methods [POK*11, SKY*12, HGP12, HGPS13]. For a detailed review of HDR deghosting algorithms, we refer the reader to a recent survey by Tursun et al. [TAEE15].

Subjective Deghosting Evaluation. Recently, subjective evaluation of HDR deghosting algorithms have begun to appear. Karaduzovic et al. [HTM13] conducted the first subjective experiment of this kind, in which 30 participants were asked to perform pairwise comparisons of four deghosting algorithms on nine different real-world scenes with varying complexity. In a more recent study, Tursun et al. carried out a larger subjective experiment, involving ten different scenes, to gather the preferences of sixty-three subjects via pairwise comparisons of six deghosting methods [TAEE15].

Image Quality Metrics. Image quality metrics are generally categorized into three classes, namely full-reference (FR), reduced-reference (RR), and no-reference (NR) metrics. FR metrics require a ground-truth reference image in addition to the image whose quality is to be inspected. VDP [Dal95], PSNR [TH94], VDM [Lub95], SSIM [WBSS04], VIF [SB06], FSIM [ZZMZ11] are some commonly used FR metrics. RR metrics, on the other hand, do not have a ground-truth image but employ some partial information about the reference [Bov05, LW09, SB11]. Finally, NR metrics do not require any information about the reference. They commonly employ knowledge about specific types of distortions and detect their 'signatures' in the distorted images [WR05, Chapter 5] [WB06, CCB11, LH11]. There are some NR metrics as well, which depend on natural scene statistics derived from artifact-free images [Sim05, SBC05, MMB12]. Alternatively, a number of studies learn some objective quality assessment functions from a collection of images and their subjective scores or user-selected artifacts and use them to provide NR metric scores [TJK11, HČA*12, YKKD12, LWC*13, LKPYYLDD14].

In the literature, there are some studies for building quality metrics for HDR images [RWPD10, Chapter 10] [SKMS06, KJF07, RFWB07, AMMS08, MKRH11, NMDSLC15]. These metrics operate on a pair of images with arbitrarily different dynamic ranges to which the standard quality assessment metrics cannot be directly applied. Our objective metric differs from these previous works in that, it is the first objective quality assessment metric designed to detect deghosting artifacts in HDR images. It can be considered as an RR metric as it requires the individual exposures in addition to the deghosting output.

3. Deghosting Artifacts

We studied the outputs of several deghosting algorithms for a variety of exposure sequences in order to understand what types of deghosting artifacts are produced. The following four types of artifacts were found to be the most common (Figure 2).

Blending. Blending artifacts occur when a dynamic object is

Figure 2: Typical deghosting artifacts. (a), (c) and (e) show the blending, gradient inconsistency and visual difference artifacts, respectively with problematic regions magnified in (b), (d) and (f). The images in this figure are obtained by directly merging the exposures without deghosting (a, b), by selecting the best-exposed input image as the only source for each pixel (c, d), and by incorrectly setting the weight values which causes singularities during the HDR assembly process (e, f).

blended to its background. This type of artifact is the most common for algorithms which aim to eliminate moving objects, such as Khan et al. [KAR06], but other types of algorithms were also found to exhibit blending artifacts to some extent.

Gradient inconsistency. Gradient inconsistencies occur when the HDR image contains new gradients that are absent in all of the exposures or the gradients that exist in the exposure stack are lost in the HDR image. This type of artifacts may occur due to banding (as new gradients will be created), blending (as gradients would be weakened), and structural distortions.

Visual differences. Visual differences occur when the deghosting result contains image details that cannot be produced from any of its constituent exposures due to various causes such as noise and corruption. Such differences can also be observed if a feature that exists in all input exposures is lost in the HDR image.

Dynamic range. This refers to the loss of contrast in dynamic regions of an exposure stack. This may happen if an algorithm chooses a single reference exposure for a dynamic region instead of using information from multiple exposures.

We note that not all artifacts are mutually exclusive and a single problematic region may contain multiple types of artifacts. However, taken together, they explain the majority of the problems in deghosting outputs.

4. An Objective Deghosting Metric

The inputs to our metric are the input exposures with accompanying exposure time and camera response information as well as the deghosted HDR image. The outputs are the distortion maps that show the location and the magnitude of the blending, gradient inconsistency, and visual difference artifacts. As for the dynamic range metric, a single scalar is produced which measures the dynamic range in the dynamic regions.

In all our metrics, we assume that the input exposures are aligned either by use of a static camera or by applying a global alignment algorithm such as median threshold bitmaps [War03]. Without this assumption, we would have to align the images within our metric, which would make our results dependent on the quality of the alignment algorithm used.

4.1. Blending Metric

Blending artifacts occur when two or more pixels that have different irradiance values are combined to produce an HDR pixel (Figure 2 (a-b)). The presence of this artifact can be detected if one knows the weights, w, used during the HDR creating process as these weights determine how much each pixel contributes to the final result:

$$I(p) = k \sum_{n=1}^{N} w_n(p) E_n(p). \quad (1)$$

Here, I is the HDR image, p represents the pixel index and k represents a normalization constant[†]. E_n is the nth exposure transferred into irradiance domain by dividing with the exposure time and applying the sensor-specific inverse camera response function.

The essence of the blending metric is the detection of pixels which differ in irradiance but are assigned high weights. However, these weights are typically unknown and must first be estimated.

4.1.1. Weight Estimation

Given the set of input images $\mathcal{L}$ and the HDR image I, the actual weights w_n used to reconstruct I are unknown; therefore, their estimates $\tilde{w}_n$ are obtained with the following non-negative least squares estimation:

$$\tilde{\mathbf{w}}'(p) = \arg\min_{\alpha} \|I(p) - D(p)\cdot\alpha\|_2^2, \quad \alpha \geq 0, \tag{2}$$

where $\tilde{\mathbf{w}}'(p) = \left[\tilde{w}'_1(p)\,\tilde{w}'_2(p)\ldots\tilde{w}'_n(p)\right]^\intercal$ is the vector of non-negative weight estimates for pixel p and $D(p)$ is $3\times N$ dictionary matrix of irradiance vectors from each irradiance map:

$$D(p) = \left[E_1(p) \mid E_2(p) \mid \ldots \mid E_N(p)\right]. \tag{3}$$

To comply with Equation 1, we normalize the weights to obtain the final weights that will be used in the blending metric: $\tilde{\mathbf{w}}(p) = [\tilde{w}_1(p)\,\tilde{w}_2(p)\ldots\tilde{w}_n(p)]^\intercal$:

$$\tilde{w}_n(p) = \tilde{w}'_n(p) \Big/ \sum_{n=1}^{N} \tilde{w}'_n(p). \tag{4}$$

For the least squares estimation, we use Lawson-Hanson [LH74] algorithm, which recovers the non-negative HDR reconstruction weights via ℓ_1-minimization [FK14]. This eliminates the need for a regularization term in the weight estimation. We validate the accuracy of our weight estimation scheme in Appendix B.

4.2. The Metric

As discussed earlier, blending artifacts occur when two irradiance values that are different from each other are simultaneously given high weights. We capture this phenomenon using the following metric:

$$Q_B(p) = \sum_{n=1}^{N-1}\sum_{m=n+1}^{N}\Big(\frac{\tilde{w}_n(p)+\tilde{w}_m(p)}{2} g(\tilde{w}_n(p),\tilde{w}_m(p)) \; h'(E_n(p),E_m(p))W_{n,m}(p)\Big), \tag{5}$$

where the function g computes the similarity of its inputs:

$$g(\tilde{w}_n(p),\tilde{w}_m(p)) = 1 - |\tilde{w}_n(p) - \tilde{w}_m(p)|. \tag{6}$$

The function h' returns the normalized Euclidean distance between two input irradiance vectors if this distance is large and 0 otherwise:

$$h'(E_n,E_m) = \begin{cases} 0, & \text{if } h(E_n,E_m) \leq \tau \\ h(E_n,E_m), & \text{otherwise,} \end{cases} \tag{7}$$

[†] The normalization is performed to make the mean irradiance of the HDR image equal to that of the middle exposure. It serves to simplify the comparisons between pixel values.

$$h(E_n(p),E_m(p)) = \frac{\|E_n(p) - E_m(p)\|_2}{\|M_{n,m}(p)\|_2}, \tag{8}$$

$$M_{n,m}(p) = \begin{bmatrix} \max\{E_n^r(p),E_m^r(p)\} \\ \max\{E_n^g(p),E_m^g(p)\} \\ \max\{E_n^b(p),E_m^b(p)\} \end{bmatrix}. \tag{9}$$

Here, τ represents a tolerance threshold. We assume that if the input irradiances are similar, their blending will not cause visible blending artifacts. We have experimented with various values of τ and found that $\tau = 0.30$ gives the highest correlation with subjective preferences and used this value for all results in this paper.

Finally, $W_{n,m}(p)$ represents the joint well-exposedness of a pixel p for exposures n and m. We use a broad-hat function to represent well-exposedness:

$$w_{BH} = 1 - (2x-1)^{32}, \tag{10}$$

$$W_{n,m}(p) = w_{BH}(L_n(p))w_{BH}(L_m(p)). \tag{11}$$

$W_{n,m}(p)$ attenuates the blending metric output for pixel p when one or both of the input pixels are under- or over-exposed. In summary, the blending metric detects those pixels in the HDR image that are created from merging well-exposed irradiance values that differ by at least 30%. The greater the magnitude of this difference, the higher the corresponding blending map value will be.

4.3. Gradient Inconsistency Metric

We assume that an HDR image should not contain any gradients that do not exist in any of its constituent exposures. Similarly, if there is a gradient in all exposures, this gradient should exist in the HDR image as well (Figure 2 (c-d)). We capture those pixels that fail these requirements in our gradient inconsistency map, which is defined as follows:

$$Q_G(p) = \begin{cases} 0, & \text{if } Q_{G'}(p) \leq \tau \\ Q_{G'}(p), & \text{otherwise.} \end{cases} \tag{12}$$

Akin to the blending metric, this branching on the magnitude of the gradient difference is made to allow small gradient differences to be tolerated. Furthermore, as a gradient is defined by its magnitude and orientation, we developed separate metrics to measure these two properties. The gradient magnitude metric is defined as:

$$Q_{G'_{mag}}(p) = \min_n \frac{\left|\frac{\overline{\|\nabla E_n\|_2}}{\overline{\|\nabla I\|_2}}\|\nabla I(p)\|_2 - \|\nabla E_n(p)\|_2\right|}{\max\left\{\frac{\overline{\|\nabla E_n\|_2}}{\overline{\|\nabla I\|_2}}\|\nabla I(p)\|_2, \|\nabla E_n(p)\|_2\right\}}, \tag{13}$$

where ∇ is the image gradient computed by using Sobel's operator, $\overline{\|\nabla E_n\|_2}$ and $\overline{\|\nabla I\|_2}$ are the mean values of $\|\nabla E_n\|_2$ and $\|\nabla I\|_2$, respectively. Normalization with the mean values is performed to make the gradient magnitudes of the HDR image compatible with the gradient magnitudes of the individual exposures. The denominator ensures that the metric output is in the range $[0,1]$.

As for the gradient orientation, we measure the minimum angle between the directions of gradient vectors:

$$Q_{G'_{dir}}(p) = \min_n |[(\theta_I(p) - \theta_n(p) + \pi) \bmod 2\pi] - \pi| / \pi \tag{14}$$

The division by π scales the output value to the range $[0,1]$. Furthermore, gradient magnitude and orientation inconsistencies are computed in a multi-scale fashion to capture gradient differences in multiple scales. To this end, we compute an image pyramid of 5 levels by Gaussian smoothing and downsampling. We evaluate Equations 13 and 14 for each pixel of each level giving rise to 5 gradient magnitude and orientation maps. We then upsample and merge artifact maps of all levels to determine the artifact maps at the finest level. During this process, the inconsistencies in higher levels are given more weight than the inconsistencies in lower levels as they correspond to more important gradient differences. We found that assigning 4 times more weight to the errors at a higher level than its immediate lower level gives the highest correlation with subjective results. This is expected as ratio of the number of pixels between neighboring pyramid levels is also 4.

4.4. Visual Difference Metric

Besides blending and gradient artifacts, deghosting outputs may also contain certain artifacts that may be collectively termed as visual differences. These may be in the form of noise, corruptions, banding, etc. To capture these general types of artifacts, we extended the HDR-VDP-2.2 metric [NMDSLC15], which is the latest and most accurate incarnation of the HDR-VDP family, to enable comparisons between a single HDR image and multiple LDR exposures.

To make the VDP metric compatible with multiple exposures, the input images, L_n, and the HDR image, I, are scaled to set their mean irradiance values to unity. Then for each pixel p, the minimum probability obtained among each HDR-LDR pair is taken as the probability of visual difference detection for that pixel:

$$Q_V(p) = \min_n V'(I(p)/\bar{I}, L_n(p)/\overline{L_n}), \tag{15}$$

where V' is the map of the visual difference detection probability generated by the HDR-VDP-2.2. The color encoding parameter of HDR-VDP-2.2 is set to ITU-R BT.709 RGB and the pixels per degree parameter is given as 30, which is an approximate value for a computer screen with a standard resolution from a typical viewing distance. Note that, this metric is not tailored for a specific type of artifact but it reports errors for any predicted visual differences.

4.5. Dynamic Range Metric

The metrics discussed so far do not award an algorithm for producing a higher dynamic range output. However, simply producing a higher dynamic range output is also not sufficient if the output contains visual artifacts. As such, an ideal algorithm should maximize the dynamic range without producing visually disturbing artifacts. In this metric, we only compute the dynamic range in the dynamic image regions. Otherwise, if the static regions have higher dynamic range, they could mask the dynamic range in the dynamic regions. To this end, we use a simple heuristic to estimate a dynamic region bitmap:

$$DR(p) = \begin{cases} 0, & \text{if } DR'(p) \leq \tau \\ 1, & \text{otherwise,} \end{cases} \tag{16}$$

where

$$DR'(p) = \max_{c\in\{r,g,b\}, n\in\{1,\ldots,N-1\}} h(E_n^c(p), E_{n+1}^c(p)) W_{n,n+1}(p) \tag{17}$$

In Figure 3, we demonstrate the output of this heuristic for several dynamic scenes. The dynamic range, D, is then computed from the HDR image pixels where $DR(p) = 1$. A small percentage of the outliers are excluded to obtain a more stable measure:

$$Q_D = \log_{10} I(p_{99\%}) - \log_{10} I(p_{1\%}). \tag{18}$$

5. Results and Validation

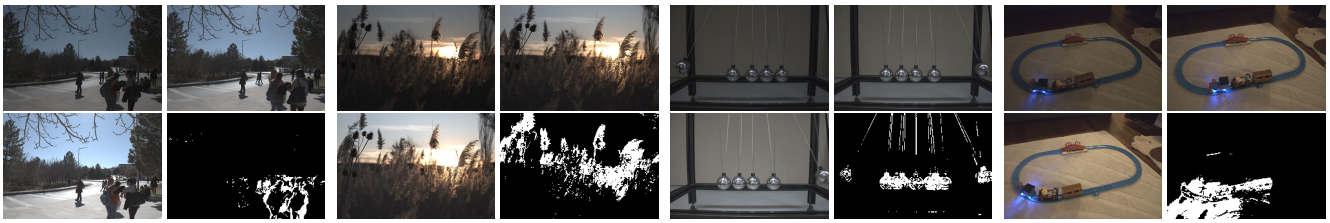
In this section, we first demonstrate the visual output of our quality metric for several deghosting algorithms applied on different exposure sets. We then compare the correlation of the proposed metric with the results of a subjective experiment. Next, we demonstrate that existing metrics that are not specialized for HDR deghosting are inadequate for detecting deghosting artifacts. Finally, we leverage our metric to illustrate an application called hybrid deghosting in which different algorithms' outputs are merged to obtain a higher quality deghosting result.

5.1. Visual Evaluation

We demonstrate three sample visual outputs of our metric in Figures 1, 4, and 5. In Figure 1 (a), Khan et al.'s [KAR06] output is shown in the bottom-left corner. The corresponding artifact maps are shown to its right as an overlay on top of the deghosting result. Here, and in all figures in this paper, blending artifacts are shown in red, gradient magnitude artifacts in green[‡], and visual difference artifacts in blue. It can be seen that our metrics detect regions that are affected by deghosting artifacts while generating only a few false positives. In Figure 1 (b), our metric primarily reports gradient magnitude inconsistencies and visual differences for the output of Hu et al.'s algorithm [HGPS13]. Comparison of this output with individual exposures reveal that there is indeed a loss of details at the back of the person as well as in the distant corridor.

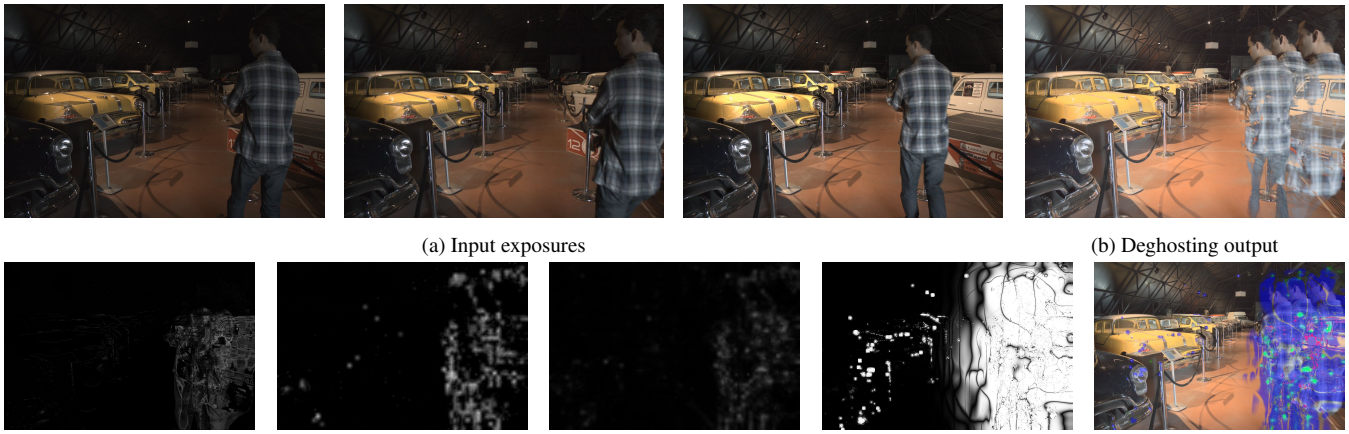
In Figure 4, we demonstrate the individual as well as the combined outputs of our metric. The input exposures are shown in (a) and the deghosting result in (b). Individual maps are shown in (c) to (f) and the combined result is shown in (g). Again, we can visually observe that most of the problematic regions in (b) are captured by our metric while generating only a few false positives.

Finally, we show that our metric can be used to detect artifacts in a larger exposure sequence as shown in Figure 5. Here, seven input exposures shown in (a) are used to obtain the deghosting result in (b) by Silk and Lang's algorithm [SL12]. The outputs of our metric are shown in (c) to (f).



(a) Scene *StreetDay* (b) Scene *Plants* (c) Scene *Pendulum* (d) Scene *ToyTrain*

Figure 3: Dynamic region bitmaps detected by our heuristic. Note that the bitmaps successfully capture dynamic objects such as people, foliage, pendulum, and a toy train while generating only a few false positives.



(a) Input exposures (b) Deghosting output

(c) Blending artifact (d) Grad. magnitude artifact (e) Grad. direction artifact (f) Visual difference (g) Combined visualization

Figure 4: Khan et al.'s [KAR06] output showing blending, gradient inconsistency and visual difference artifacts.

5.2. Validation

To understand whether these visual observations can be generalized to a larger set of images, we conducted a subjective experiment involving 10 deghosting algorithms belonging to different classes (with one of them being no-deghosting as a control condition), 16 scenes of varying characteristics, and 52 participants. The input scenes are represented in Figure 6 and their characteristics are described in Table 1. The purpose of this experiment was to collect subjective rating data and then evaluate the correlation of our metric outputs with this data. The details and the results of this subjective experiment are given in Appendix A.

As three of our metrics generate a distortion map whereas participants assign a single quality score, we computed a global quality score from each distortion map, i, as follows:

$$Q_i = -\sum_{p \in P} Q_i(p)/|P|, \tag{19}$$

where each quality metric is substituted for i and $|P|$ represents the number of image pixels. The result is negated to yield more negative scores for worse results. As for the dynamic range metric, we directly used the dynamic range in dynamic regions (Equation 18). We computed the Pearson and Spearman correlation coefficients between these scores and the aggregate ratings obtained from the subjective experiment. As both correlations were very similar, here we only report the Pearson correlation results (full results are given in supplementary materials).

Furthermore, we experimented with two modes in which our metric scores are computed. In the first mode, we directly computed the metric outputs as defined by Equation 19. In the second mode, we first computed a visual saliency map using Itti et al.'s model [IKN98] using the deghosting output as the input image to this model[§]. Using the saliency maps helps us to take into account the visually important image pixels in estimating the metric scores. We then computed the weighted average of the saliency map with the distortion maps:

$$Q_i^s = -\sum_{p \in P} \sqrt{Q_i(p)S(p)}/|P|, \tag{20}$$

where $S(p)$ represents the saliency map value. We found that using saliency only improves correlations for the gradient metrics while

[‡] In overlay visualizations, we only show gradient direction artifacts for clarity. As we show later, they have a significant overlap with gradient magnitude artifacts.

[§] We used the implementation by Jonathan Harel: A Saliency Implementation in MATLAB: http://www.klab.caltech.edu/~harel/share/simpsal.

(a) Input exposures

(b) Deghosting Output (c) Blending Map (d) Grad. Magnitude Inconsistency (e) Grad. Direction Inconsistency (f) Visual Difference Map

Figure 5: Metric outputs for scene *ToyTrain* using 7 exposures with the algorithm of Silk and Lang [SL12].

Figure 6: The representative images of the input scenes used in the experiment.

having a slightly adverse effect for the blending metric. This could be because the image regions that contain blending artifacts are not that salient due to the softened appearance of objects. The dynamic range metric produces a single scalar value so it is inappropriate to be used in this equation. Finally, the extended VDP metric is a self-contained model which is already based on visual perception, so we did not try to combine it with the saliency map.

As can be seen in Table 2, all our metrics are positively correlated with the subjective experiment results. The aggregate correlations of all metrics except $Q^s_{G'_{mag}}$ are above 0.50 with the extended VDP being the highest (0.68) followed by blending (0.66), gradient direction (0.62), dynamic range (0.59), and gradient magnitude (0.49) metrics. As for the consistency across scenes, blending metric yields the lowest standard deviation (0.14) followed by VDP (0.21), dynamic range (0.21), gradient direction (0.23), and gradient magnitude (0.31).

Next, we evaluated the correlation of the individual metrics with each other for all scene-algorithm combinations. This was performed to understand the degree of the overlap between the outputs produced by different metrics. Table 3 summarize these results. According to this table, there is a high degree of correlation between the gradient magnitude and direction metrics. As the latter has a higher mean correlation and lower standard deviation, we decided to drop the gradient magnitude metric for improved reliability.

The analyses of the individual scenes reveal that these metrics support each other. In some scenes, they all have high correlations with subjective ratings (scenes *Pendulum* and *ToyTrain*) while in others one or two metrics make up for the low correlation of other metrics. For example in scene *StreetNight* the Q_B and Q_D have high correlations (0.68 and 0.66 respectively) whereas $Q^s_{G'_{dir}}$, $Q^s_{G'_{mag}}$, and Q_V have low correlations (0.06, −0.15, and 0.24). This suggests that combining these metrics to yield a single quality score could in fact produce a higher correlation than all metrics taken alone. To test this hypothesis, we computed a unified deghosting quality metric (UDQM) score Q_U for each deghosting result as a weighted sum of the individual metrics:

$$Q_U = [Q_B \; Q^s_{G'_{dir}} \; Q_V \; Q_D \; 1] \cdot [w_B \; w_G \; w_V \; w_D \; w_c]^\mathsf{T}, \tag{21}$$

where w_i indicates the weight of the quality score i and w_c is the weight of the constant term, which will be used during the regression analysis.

In order to find the best weights, we used adaptive simulated annealing [I*96] for maximizing the average Pearson correlation between Q_U and subjective ratings. For simulated annealing, the initial temperature is set to $T_0 = 100$ and the following annealing schedule is used:

$$T = 0.95^k T_0, \tag{22}$$

where k is the annealing parameter. The probability of acceptance function is:

$$P = \frac{1}{1 + \exp\left(\dfrac{\Delta E}{\max\limits_i(T_i)}\right)}, \tag{23}$$

Table 1: Properties of the scenes used in the experiment.

Scene	Properties
StreetNight	Walking person with some distance to the camera, outdoors, night, motion blur
Museum1	Walking person close to the camera, indoors
Museum2	Mostly static high contrast scene with a person in one exposure
Museum3	Movement of a close person across a high contrast distant door
StreetDay	Multiple moving people, outdoors, day
Plants	Complex motion of plants across sunset
Exit	Person exiting from dark indoors into sunny outdoors
Pendulum	Complex motion of objects with specular highlights
Cars	Fast motion of cars across sunset
Pedestrians	Slow motion of pedestrians across sunset
Fountain	Complex motion of water fountain with crowded background
Yard	Multiple people walking away from the camera
Turtles	Relatively still water with changing reflections
Building	Closeby person walking in front of large windows
Flames	Complex motion of flames
ToyTrain	Movement of a toy train along tracks with flashing lights

where ΔE is the difference between the present and past values of the energy function and T_i is the current temperature of component i. No upper and lower bounds are imposed on the metric weights; however, the resulting value of vector $\mathbf{w}$ is normalized to have a length of 1. The normalization has no effect on the value of the energy function (i.e. average Pearson correlation) and it is only performed to have metric weights with comparable scales in different scenes for our analysis.

We performed leave-one-out cross-validation (LOOCV) to help validate if the weights learned from all images except one is a good indicator for the left-out image. The individual weights estimated by using adaptive simulated annealing after leaving out each one of the images is given in Table 4. We observed that the estimated weights are consistent with each other. The mean LOOCV value is found to be 0.77 with a standard deviation of 0.14. This high correlation and low standard deviation suggest that the weights learned from a subset of the scenes (training) can be used to estimate the quality of outputs that are not part of the dataset (testing). Therefore, to produce our final set of weights we have included all scenes in the optimization, which resulted in the weights shown in Table 5.

As an additional validation of the proposed weights, we used them to compute a UDQM score for a set of images used in an independent subjective experiment performed by Tursun et al. [TAEE15]. Our goal was to investigate how well the computed scores correlate with the subjective responses from an entirely independent dataset. The experiment was a pairwise comparison experiment involving 7 deghosting algorithms and 10 scenes. We computed the number of times an algorithm was selected over another one as the rating of that algorithm (as given in Tables 3 and 4 of Tursun et al. [TAEE15]), and then correlated the UDQM scores with those ratings. The resulting correlations are given in Table 6. As can be seen from this table, we obtained a mean Pearson correlation of 0.65 with 0.20 standard deviation. Although the mean correlation was somewhat smaller than the correlation of the experiment conducted in this paper, it is still high enough to be predictive of the algorithms' performance.

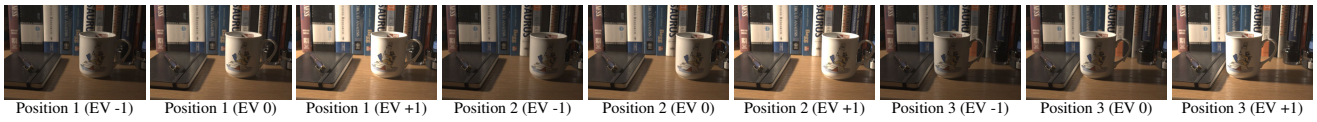
Finally, we performed an analysis to understand how well an individual participant's responses correlate with the mean responses of all participants except herself. As we had 52 participants in our experiment, this produced 52 correlation scores with the mean score being 0.75 and the standard deviation 0.22. Given that our mean UDQM correlation for the test set was 0.77 with a standard deviation of 0.14, it is not unreasonable to assume that our combined metric's responses resemble that of an average observer.

Table 2: Pearson correlations of the individual metrics with subjective ratings.

Scene	Q_B	$Q^s_{G'_{dir}}$	$Q^s_{G'_{mag}}$	Q_D	Q_V
StreetNight	0.68	0.06	-0.22	0.66	0.24
Museum1	0.91	0.87	0.85	0.65	0.65
Museum2	0.66	0.56	0.45	0.57	0.76
Museum3	0.70	0.67	0.18	0.28	0.41
StreetDay	0.85	0.55	0.34	0.62	0.85
Plants	0.57	0.76	0.64	0.63	0.82
Exit	0.65	0.32	0.07	0.49	0.25
Pendulum	0.82	0.77	0.59	0.73	0.83
Cars	0.60	0.63	0.79	0.51	0.58
Pedestrians	0.57	0.39	0.33	0.84	0.78
Fountain	0.65	0.82	0.43	0.20	0.70
Yard	0.55	0.39	0.27	0.51	0.72
Turtles	0.51	0.81	0.84	0.93	0.91
Building	0.65	0.73	0.81	0.52	0.81
Flames	0.39	0.76	0.66	0.40	0.67
ToyTrain	0.78	0.79	0.76	0.94	0.85
Average	0.66	0.62	0.49	0.59	0.68
Std. Dev.	0.14	0.23	0.31	0.21	0.21

Table 3: Pearson correlation coefficients between absolute metric scores in all scene-algorithm combinations.

	Q_B	$Q^s_{G'_{dir}}$	$Q^s_{G'_{mag}}$	Q_D	Q_V
Q_B	1.00	0.60	0.45	-0.51	0.53
$Q^s_{G'_{dir}}$	0.60	1.00	**0.83**	-0.24	0.64
$Q^s_{G'_{mag}}$	0.45	**0.83**	1.00	-0.21	0.66
Q_D	-0.51	-0.24	-0.21	1.00	-0.51
Q_V	0.53	0.64	0.66	-0.51	1.00



(a) Input exposures for three different positions of the mug

(b) Ground truth 1 (c) Ground truth 2 (d) Ground truth 3 (e) Output HDR (f) Difference Map (g) Our metric results

Figure 7: Comparison of our metric result (g) with the absolute difference (f) between the output of Silk and Lang [SL12] (e) and ground truth HDR (c). The input exposures with EV -1, 0 and +1 are shown in (a). For each position of the moving object, the ground truth HDR image obtained by using the input exposures is given in (b-d).

Table 4: Best metric weights found in LOOCV and their corresponding Pearson correlations in training and testing scenes. The scene which is left out is given in the first column.

Scene	w_B	w_G	w_D	w_V	w_c	Train	Test
StreetNight	0.548	0.722	0.025	0.035	0.420	0.79	0.63
Museum1	0.510	0.413	0.021	0.051	0.753	0.77	0.87
Museum2	0.495	0.822	0.031	0.045	-0.276	0.77	0.88
Museum3	0.449	0.891	0.035	0.048	-0.036	0.79	0.59
StreetDay	0.794	0.600	0.023	0.044	-0.090	0.77	0.86
Plants	0.631	0.491	0.018	0.036	-0.600	0.78	0.78
Exit	0.415	0.881	0.030	0.036	0.223	0.80	0.43
Pendulum	0.628	0.705	0.025	0.036	0.325	0.77	0.89
Cars	0.547	0.780	0.028	0.043	0.299	0.78	0.75
Pedestrians	0.539	0.810	0.023	0.028	0.230	0.78	0.82
Fountain	0.086	0.813	0.035	0.037	-0.573	0.78	0.64
Yard	0.425	0.898	0.033	0.031	0.106	0.78	0.72
Turtles	0.695	0.717	0.024	0.027	-0.039	0.78	0.80
Building	0.424	0.848	0.032	0.037	-0.315	0.78	0.80
Flames	0.248	0.934	0.041	0.072	0.244	0.77	0.90
ToyTrain	0.411	0.905	0.034	0.056	0.094	0.77	0.95
Average	0.490	0.764	0.029	0.041	0.048	0.78	0.77
Std. Dev.	0.169	0.151	0.006	0.011	0.361	0.01	0.14

Table 5: Best metric weights obtained using adaptive simulated annealing [I*96].

w_B	w_G	w_D	w_V	w_c
0.427	0.811	0.029	0.037	0.397

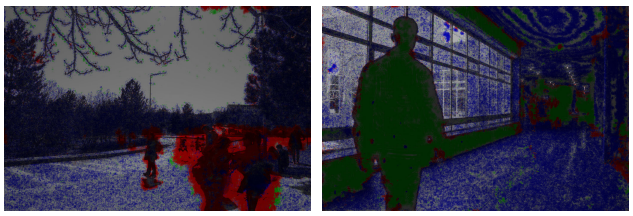
5.3. Comparison with Other Quality Metrics

In this section, we demonstrate the outputs of the two of the potentially relevant quality metrics, namely the dynamic range independent quality metric (DRIM) [AMMS08] and a blind deblurring quality metric [LWC*13]. The former metric is selected as it can compare images with different dynamic ranges and the latter is selected because deblurring artifacts may be somewhat similar to deghosting artifacts.

The distortion map produced by the DRIM for the images shown in Figure 1 are shown in Figure 8. To compute these results, the middle exposure of the exposure sequence is compared against the deghosting output. In DRIM outputs, red indicates reversal of contrast, green loss of contrast, and blue amplification of contrast. We can see that although the DRIM metric detects problematic regions it also generates many false positives which correspond to mostly static parts of the scenes (blue regions). As such, the DRIM metric is not suitable to be used for detecting deghosting artifacts.

Table 6: Pearson correlation coefficients of Q_U from the dataset of Tursun et al. [TAEE15] using the weights in Table 5.

Cafe	Candles	FastCars	Flag	Gallery1	Gallery2
0.90	0.67	0.64	0.27	0.64	0.61
LibrarySide	**Shop1**	**Shop2**	**WalkingPeople**	**Average**	**Std. Dev.**
0.95	0.66	0.43	0.71	0.65	0.20



(a) DRIM (Khan et al. [KAR06]) (b) DRIM (Hu et al. [HGPS13])

Figure 8: The DRIM [AMMS08] outputs for two sample scenes. In DRIM, blue represents amplification of contrast, green loss of contrast, and red reversal of contrast. Our metric results were shown in Figure 1.

The blind deblurring metric of Liu et al. [LWC*13] produces a single quality score instead of a distortion map. In order to calculate this quality score, all input HDR images are tone mapped with the tone mapping operator of Reinhard et al. [RSSF02] and gamma corrected ($\gamma = 2.2$). For no-reference ringing detection, the original metric uses the blurry input image in addition to the deblurred images. As we do not have a blurry input image in our scenario, instead we used the image obtained by directly merging the exposures without applying deghosting. We generated the quality scores for all scenes used in the subjective experiment and computed the correlations with participant ratings. As shown in Table 7, this met-

Table 7: Pearson correlations for Liu et al.'s [LWC*13] deblurring metric with the subjective experiment.

Exit	Pendulum	Cars	Pedestrians	Fountain	Yard
0.16	-0.01	0.30	0.64	0.11	0.50
Turtles	**StreetNight**	**Building**	**Flames**	**ToyTrain**	**Museum1**
0.76	0.08	-0.24	-0.22	0.24	0.22
Museum2	**Museum3**	**StreetDay**	**Plants**	**Avg.**	**Std.**
0.36	-0.21	0.33	0.58	0.23	0.30

ric has a low correlation with subjective ratings and therefore is not suitable to be used as a deghosting quality metric.

5.4. Comparison with Ground Truth

In order to show how well our metric captures the visual artifacts, we prepared ground truth HDR images and computed the absolute difference map between the deghosting result and one of the ground truth HDR images for comparison with our metric output. For this purpose, we captured three exposures with EV -1, 0 and $+1$ for three different positions of an object in Figure 7 (a). After this process, we created a ground truth HDR image for each position of the object in the scene (b, c, d). Then we applied the deghosting algorithm of Silk and Lang [SL12] to create an HDR image using EV -1, 0 and $+1$ from object positions 1, 2 and 3, respectively (e). We chose the ground truth 2 as the reference for creating the difference map since it was visually the most similar ground truth image to the deghosting result. The minimum and 99^{th} percentile absolute radiance differences were then scaled between 0 (black) and 1 (white) for visualization in the difference map (f). Our metric output is shown in (g). It can be observed that the highest responses of our metrics correspond to the pixels with largest absolute differences. This figure also shows that the visual difference metric is more sensitive to the artifacts over a larger area while blending and gradient inconsistency metrics give more localized responses.

5.5. Application: Hybrid Deghosting

Finally, we demonstrate that by leveraging our metric one can combine the outputs of multiple deghosting algorithms to obtain a higher quality deghosting result. To perform hybrid deghosting, we first calculated the UDQM map for each one of the input HDR images. We then extracted the binary mask of pixels from each image, which will contribute to the hybrid HDR. A pixel is marked in the corresponding binary mask if it has the best UDQM value among all the input HDR images. Before combining the images by using Laplacian blending according to these masks, we applied histogram matching to ensure that the irradiance distributions of the two images are similar.

Two sample outputs are provided in Figures 9 and 10. In Figure 9, we combined Lee et al.'s [LLM14] output with a simple deghosting approach, which involves selecting the middle exposure in all dynamic regions. In Figure 10, on the other hand, Hu et al. [HGPS13]'s and Lee et al.'s [LLM14] results are combined using their distortion maps. In both figures, the combined outputs have fewer artifacts than the individual results.

It should be noted, however, that hybrid deghosting is not suitable to be used for all pairs of algorithms, especially for those which select different reference exposures. In such cases, multiple copies of the same object could appear in the combined result. An example of this limitation is shown in Figure 11, where the inconsistency of the HDR images due to the difference in object positions result in a hybrid HDR image with noticeable blending and gradient inconsistency artifacts.

6. Conclusion and Future Work

In this paper, we described an HDR deghosting quality metric, comprised of individual metrics tuned for different types of artifacts, and showed that they correlate well with visual observations and subjective preferences. The first application that our metric enables is automatic quality evaluation of deghosting algorithms. By using our metric, one can often avoid comprehensive subjective experiments which are both tedious and may become outdated as new algorithms are proposed. Secondly, the proposed metrics can be used to optimize parameter selection for deghosting algorithms. For instance, a deghosting method of choice may be run in batch-processing mode and allowed to explore a parameter space until it finds the combination that gives the least distortions as computed by our metric. Our metric can also be used to rapidly assess the strengths and weaknesses of different algorithms for different images sets and ultimately design improved algorithms. Finally, as demonstrated in the last section, our metric also enables a novel application called hybrid deghosting, whereby multiple deghosting results are combined to obtain a higher quality result.

Using our metric to provide feedback for artifact removal is a natural research direction to improve the outcomes of the HDR deghosting algorithms. As a more ambitious future work, we plan to explore a no-reference quality metric which takes only the deghosted HDR image as input, but not its constituent exposures.

A. Subjective Experiment

In order to validate the compatibility of our metrics with subjective preferences, we conducted a subjective experiment involving 10 deghosting algorithms and 16 scenes. The selected algorithms were: Grosch [Gro06], Khan et al. [KAR06], Sen et al. [SKY*12], Silk and Lang's [SL12] fluid-motion (FM) and pairwise-downweighting (PWD) methods, Srikantha et al. [SSM12], Hu et al. [HGPS13], Lee et al. [LLM14], a simple deghosting approach based on using a single reference image in all dynamic regions, and no deghosting as a control condition. These algorithms were selected as representatives of different types of deghosting approaches. They range from simple point operations to sophisticated computer vision algorithms.

We acquired the source codes and/or executables of each of these algorithms from the original authors except for Khan et al.'s method [KAR06], which we implemented ourselves. We also slightly modified Silk and Lang's algorithm [SL12] to prevent it producing NaN values due to mathematical singularities.

As for the input scenes, we created a dataset of diverse characteristics with both indoor and outdoor environments and motion

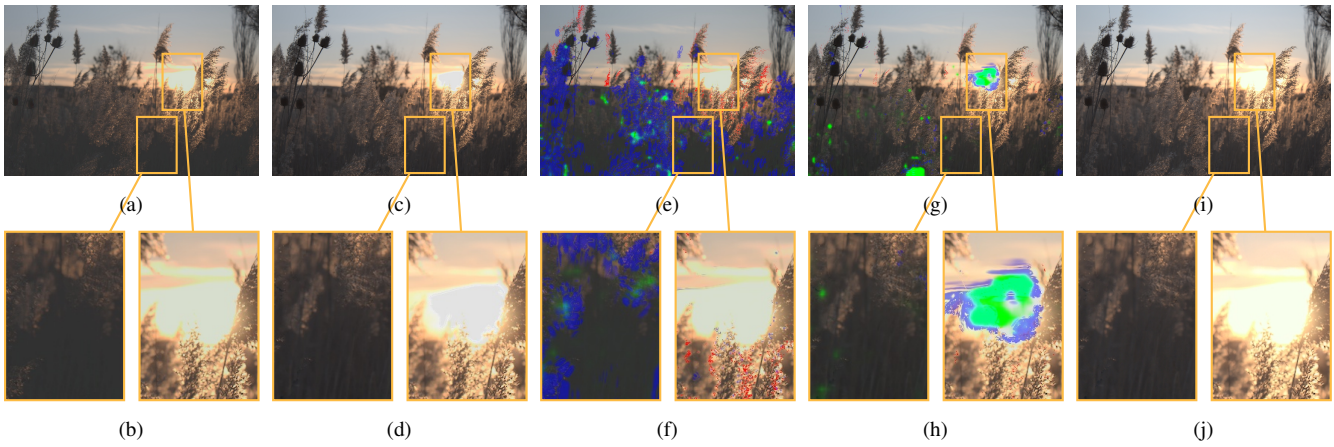


Figure 9: The outputs of Lee et al. [LLM14] (a-b) and simple deghosting approach (c-d), our results for Lee et al. [LLM14] (e-f) and simple deghosting approach, and our hybrid deghosting result (i-j).

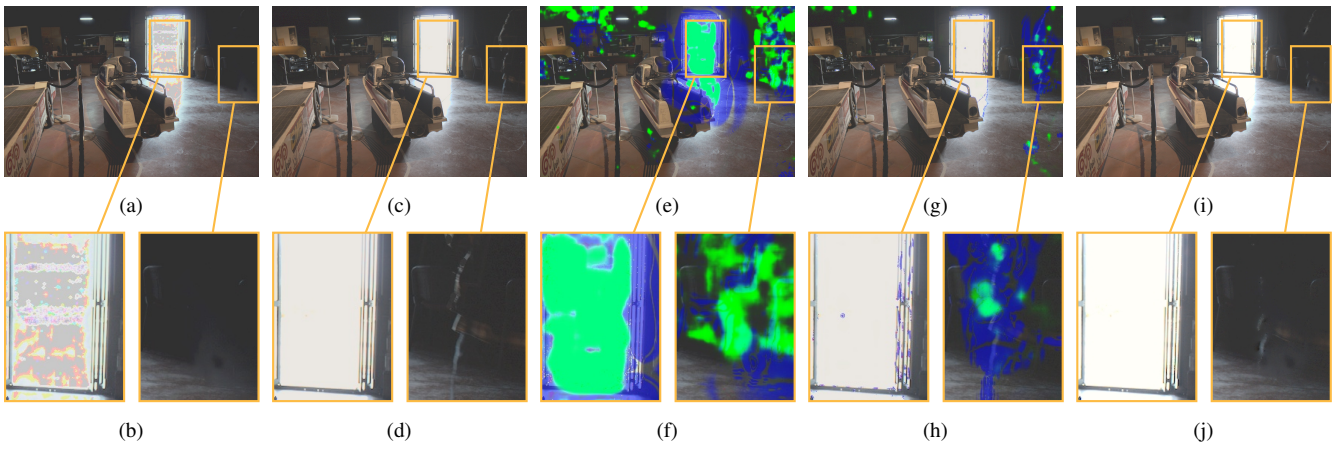


Figure 10: The outputs of Hu et al. [HGPS13] (a-b) and Lee et al. [LLM14] (c-d), our results for Hu et al. [HGPS13] (e-f) and Lee et al. [LLM14], and our hybrid deghosting result (i-j).

Figure 11: The hybrid HDR image in (c) is obtained from two HDR outputs of Srikantha et al. [SSM12] (a) and Lee et al. [LLM14] (b), which are inconsistent with each other in terms of object positions.

patterns with varying complexities. Each scene was captured using 3 exposures (-1, 0, +1 EV) with a Canon DSLR camera in RAW format. These scenes are depicted in Figure 6 and their properties are summarized in Table 1.

Due to a large number of algorithm/scene combinations, we performed an online rating experiment. The experiment started by a warm-up trial to familiarize the participants with the web-interface. The interface consisted of thumbnails of three individual exposures on the left and ten deghosting results on the right. The selected exposure and the deghosting result were shown in higher resolution side-by-side in the middle of the screen. Hovering the mouse over these images brought a zoomed-in view of the region under the cursor to allow detailed analysis.

The participants' task was to give a rating between 0 and 100 for each deghosting result by setting the sliders below each thumbnail. To facilitate these ratings, we marked 0, 25, 50, 75, and 100 values as 'very bad', 'bad', 'medium', 'good', and 'very good' quality, respectively. However, the participants could assign any rating between 0 and 100 with a step size of 5 (e.g. a good image could be rated as 75 whereas a slightly better one as 80). After each trial, the participants viewed the next set of images by pressing the 'Next' button at the bottom of the page. The experiment took between 30 to 40 minutes for each participant. We discarded partially completed experiments to have equal number of ratings for all condi-

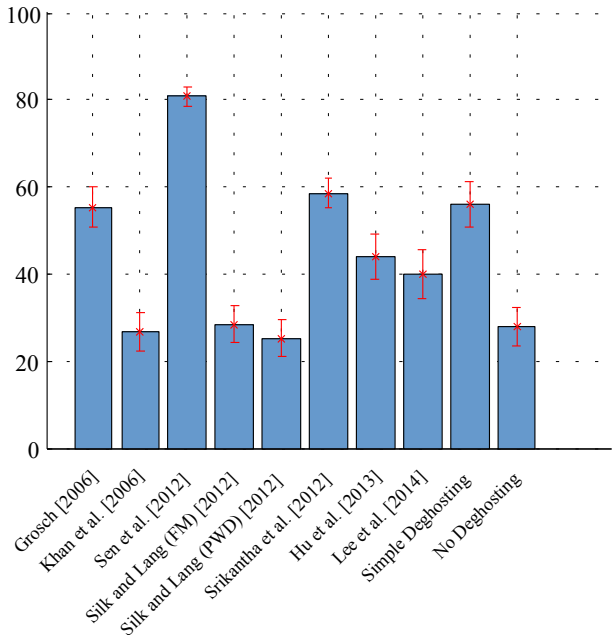


Figure 12: Average ratings of the compared algorithms in the subjective experiment. The red lines indicate the standard error.

tions. In total, 52 participants finished the experiment in its entirety and their mean ratings are shown in Figure 12.

B. Validation of Weight Estimation

We evaluated our weight estimation scheme to determine how well it can recover the actual weights used during the HDR creation. To this end, we created three HDR images using three different weighting functions, namely triangular (w_T), broad-hat (w_{BH}), and Gaussian (w_G) functions. These functions were defined as follows:

$$w_T(x) = 1 - |2(x-0.5)|, \tag{24}$$

$$w_{BH}(x) = 1 - (2x-1)^{32}, \tag{25}$$

$$w_G(x) = \exp(-25(x-0.5)^2), \tag{26}$$

where x is the normalized pixel value. The correlation between the estimated weights and the actual weights were found to be moderate (in the range $[0.3, 0.6]$ for different weighting functions). This is expected as there may be many different set of weights which produce the same HDR pixel from a set of LDR pixels. The critical requirement should be that the final HDR images that are obtained with the actual weights and the estimated weights should be highly correlated. These correlation values were found to be greater than 0.99 for all three weighting functions.

We performed a follow-up experiment to evaluate the similarity of the blending maps obtained using the actual weights and estimated weights. As shown in Figure 13, an artificial exposure sequence was created, which contained a single moving object. We combined these exposures using a triangular weighting function and without ghost removal to obtain the results shown in (b). We then computed the blending map using these *known* weights as shown in (c). Next, we estimated the weights from the HDR image and the input exposures using our weight estimation scheme. The blending map computed from these estimated weights are shown in (d). These two blending maps are visually very similar and have a Pearson correlation coefficient of 0.96 when 3 input exposures are used. We then used 7 input exposures to understand how well our weight estimation scheme can deal with a larger number of exposures and found a correlation of 0.88 between the maps computed from the actual and estimated weights. The HDR image obtained from the estimated weights shown in (e) is also highly correlated with the HDR image reconstructed by using the actual weights (b).

These experiments support that the estimated weights can be reliably used to detect blending artifacts in most cases. The only limitation of our scheme is that if two irradiance values are linearly dependent to each other. In that case, the deghosting algorithm may have assigned non-zero weights to two different exposures with different but linearly dependent irradiance vectors, producing a ghost artifact. This type of artifact is not detected by our blending metric as our ℓ_1-minimization would assign a non-zero weight only to the higher irradiance pixel value. However, in practice, the probability of having linearly dependent color vectors in the same pixel of two different exposures is low due to noise.

Another implication of our weight estimation scheme is that if the number of exposures is larger than the number of color channels, the non-negative least squares estimation algorithm performs a sparse recovery of the weights. As a result, the number of non-zero elements in α is limited to 3 in our case, even if the number of input exposures is greater than 3. However, this does not preclude detecting blending artifacts as shown in the bottom row of Figure 13. It simply means that we can detect the blending artifacts caused by the three highest irradiance pixels.

References

[AMMS08] AYDIN T. O., MANTIUK R., MYSZKOWSKI K., SEIDEL H.-P.: Dynamic range independent image quality assessment. *ACM Transactions on Graphics 27*, 3 (Aug. 2008), 69:1–69:10. 2, 9

[Bov05] BOVIK A. C.: *Handbook of Image and Video Processing (Communications, Networking and Multimedia)*. Academic Press, Inc., Orlando, FL, USA, 2005. 2

[CCB11] CHEN C., CHEN W., BLOOM J. A.: A universal reference-free blurriness measure. *Proc. SPIE 7867* (2011), 78670B–78670B–14. 2

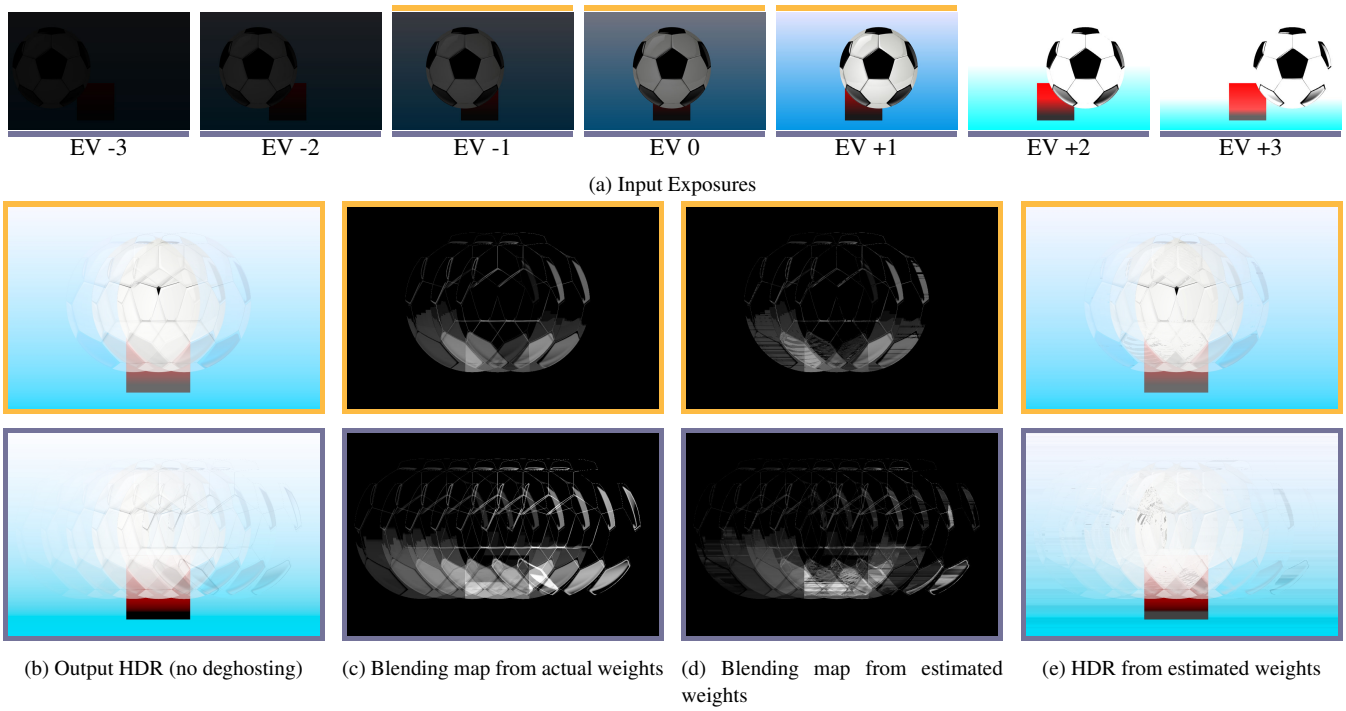
[Dal95] DALY S.: The visible differences predictor: An algorithm for the assessment of image fidelity. In *Digital Images and Human Vision*, Watson A. B., (Ed.). MIT Press, 1995, pp. 179–206. 2

[DM97] DEBEVEC P. E., MALIK J.: Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH 97 Conference Proceedings* (August 1997), pp. 369–378. 1

[FK14] FOUCART S., KOSLICKI D.: Sparse recovery by means of nonnegative least squares. *IEEE Signal Processing Letters 21*, 4 (April 2014), 498–502. 4

[GGC*09] GALLO O., GELFAND N., CHEN W.-C., TICO M., PULLI K.: Artifact-free high dynamic range imaging. In *IEEE International Conference on Computational Photography (ICCP)* (2009), IEEE, pp. 1–7. 2

[GKTT13] GRANADOS M., KIM K. I., TOMPKIN J., THEOBALT C.: Automatic noise modeling for ghost-free hdr reconstruction. *ACM Transactions on Graphics (TOG) 32*, 6 (2013), 201. 2



(a) Input Exposures

(b) Output HDR (no deghosting) (c) Blending map from actual weights (d) Blending map from estimated weights (e) HDR from estimated weights

Figure 13: Blending metric outputs for 3 and 7 exposures using the actual and estimated HDR reconstruction weights on a set of synthetic input images. It can be observed that the blending metric maps in (c) and (d) are consistent with the visible artifacts in (b).

[Gro06] GROSCH T.: Fast and robust high dynamic range image generation with camera and object movement. In *Proc. of Vision Modeling and Visualization* (2006), pp. 277–284. 10

[HČA*12] HERZOG R., ČADÍK M., AYDIN T. O., KIM K. I., MYSZKOWSKI K., SEIDEL H.-P.: NoRM: no-reference image quality metric for realistic image synthesis. *Computer Graphics Forum 31*, 2 (2012), 545–554. 2

[HDW14] HAFNER D., DEMETZ O., WEICKERT J.: Simultaneous hdr and optic flow computation. In *22nd International Conference on Pattern Recognition (ICPR 2014)* (2014), IEEE. 2

[HGP12] HU J., GALLO O., PULLI K.: Exposure stacks of live scenes with hand-held cameras. In *European Conference on Computer Vision*. Springer, 2012, pp. 499–512. 2

[HGPS13] HU J., GALLO O., PULLI K., SUN X.: Hdr deghosting: How to deal with saturation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2013). 1, 2, 5, 9, 10, 11

[HLL*11] HEO Y. S., LEE K. M., LEE S. U., MOON Y., CHA J.: Ghost-free high dynamic range imaging. In *Asian Conference on Computer Vision (ACCV)*. Springer, 2011, pp. 486–500. 2

[HT13] HADZIABDIC K. K., TELALOVIĆ J. H.: Report: State-of-the-art de-ghosting algorithms for high dynamic range imaging. *SouthEast Europe Journal of Soft Computing 2*, 2 (2013). 2

[HTM13] HADZIABDIC K. K., TELALOVIĆ J. H., MANTIUK R.: Comparison of deghosting algorithms for multi-exposure high dynamic range imaging. In *Spring Conference on Computer Graphics* (2013), ACM, pp. 21–28. 2

[HTM14] HADZIABDIC K. K., TELALOVIĆ J. H., MANTIUK R.: Expert evaluation of deghosting algorithms for multi-exposure high dynamic range imaging. In *HDRi2014 - Second International Conference and SME Workshop on HDR imaging* (Sarajevo, Bosnia and Herzegovina, April 2014). 2

[I*96] INGBER L., ET AL.: Adaptive simulated annealing (ASA): Lessons learned. *Control and cybernetics 25* (1996), 33–54. 7, 9

[IKN98] ITTI L., KOCH C., NIEBUR E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence 20*, 11 (1998), 1254–1259. 6

[JLW08] JACOBS K., LOSCOS C., WARD G.: Automatic high-dynamic range image generation for dynamic scenes. *IEEE Computer Graphics and Applications 28*, 2 (2008), 84–93. 2

[JO12] JINNO T., OKUDA M.: Multiple exposure fusion for high dynamic range image acquisition. *IEEE Transactions on Image Processing 21*, 1 (2012), 358–365. 2

[KAR06] KHAN E. A., AKYÜZ A. O., REINHARD E.: Ghost removal in high dynamic range images. *IEEE International Conference on Image Processing* (2006). 1, 2, 3, 5, 6, 9, 10

[KJF07] KUANG J., JOHNSON G. M., FAIRCHILD M. D.: iCAM06: A refined image appearance model for hdr image rendering. *Journal of Visual Communication and Image Representation 18*, 5 (2007), 406 – 414. Special issue on High Dynamic Range Imaging. 2

[KUWS03] KANG S. B., UYTTENDAELE M., WINDER S., SZELISKI R.: High dynamic range video. In *ACM Transactions on Graphics (TOG)* (2003), vol. 22, ACM, pp. 319–325. 2

[LC09] LIN H.-Y., CHANG W.-Z.: High dynamic range imaging for stereoscopic scene representation. In *16th IEEE International Conference on Image Processing (ICIP)* (2009), IEEE, pp. 4305–4308. 2

[LH74] LAWSON C. L., HANSON R. J.: *Solving least squares problems*, vol. 161. SIAM, 1974. 4

[LH11] LIU H., HEYNDERICKX I.: Issues in the design of a no-reference metric for perceived blur. *Proc. SPIE 7867* (2011), 78670C–78670C–8. 2

[LKPYYLDD14] LE KANG, PENG YE, YI LI, DAVID DOERMANN:

Convolutional Neural Networks for No-Reference Image Quality Assessment. In *IEEE Conf. On Computer Vision and Pattern Recognition (CVPR)* (June 2014). 2

[LLM14] LEE C., LI Y., MONGA V.: Ghost-free high dynamic range imaging via rank minimization. *IEEE Signal Processing Letters 21*, 9 (Sept 2014), 1045–1049. 10, 11

[Lub95] LUBIN J.: A visual discrimination model for imaging system design and evaluation. In *Vision Models for Target Detection and Recognition*, Peli E., (Ed.). World Scientific, 1995, pp. 245–283. 2

[LW09] LI Q., WANG Z.: Reduced-reference image quality assessment using divisive normalization-based image representation. *IEEE Journal of Selected Topics in Signal Processing 3*, 2 (2009), 202–211. 2

[LWC*13] LIU Y., WANG J., CHO S., FINKELSTEIN A., RUSINKIEWICZ S.: A no-reference metric for evaluating the quality of motion deblurring. *ACM Transactions on Graphics 32*, 6 (Nov. 2013), 175:1–175:12. 2, 9, 10

[MKRH11] 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. *ACM Transactions on Graphics 30*, 4 (July 2011), 40:1–40:14. 2

[MMB12] MITTAL A., MOORTHY A. K., BOVIK A. C.: No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing 21*, 12 (2012), 4695–4708. 2

[NMDSLC15] NARWARIA M., MANTIUK R. K., DA SILVA M. P., LE CALLET P.: HDR-VDP-2.2: A calibrated method for objective quality prediction of high-dynamic range and standard images. *Journal of Electronic Imaging 24*, 1 (2015), 010501. 2, 5

[OLK13] OH T. H., LEE J.-Y., KWEON I.-S.: High dynamic range imaging by a rank-1 constraint. In *ICIP* (2013), pp. 790–794. 2

[POK*11] PARK S.-C., OH H.-H., KWON J.-H., CHOE W., LEE S.-D.: Motion artifact-free hdr imaging under dynamic environments. In *18th IEEE International Conference on Image Processing (ICIP)* (2011), IEEE, pp. 353–356. 2

[RC11] RAMAN S., CHAUDHURI S.: Reconstruction of high contrast images for dynamic scenes. *The Visual Computer 27*, 12 (2011), 1099–1114. 2

[RFWB07] RAMANARAYANAN G., FERWERDA J., WALTER B., BALA K.: Visual equivalence: Towards a new standard for image fidelity. *ACM Transactions on Graphics 26*, 3 (July 2007). 2

[RKC09] RAMAN S., KUMAR V., CHAUDHURI S.: Blind de-ghosting for automatic multi-exposure compositing. In *ACM SIGGRAPH ASIA 2009 Posters* (2009), ACM, p. 44. 2

[RSSF02] REINHARD E., STARK M., SHIRLEY P., FERWERDA J.: Photographic tone reproduction for digital images. In *ACM Transactions on Graphics (TOG)* (2002), vol. 21, ACM, pp. 267–276. 9

[RWPD10] REINHARD E., WARD G., PATTANAIK S., DEBEVEC P.: *High Dynamic Range Imaging: Acquisition, Display and Image-Based Lighting*, second edition ed. Morgan Kaufmann, San Francisco, 2010. 2

[SB06] SHEIKH H. R., BOVIK A. C.: Image information and visual quality. *IEEE Transactions on Image Processing 15*, 2 (2006), 430–444. 2

[SB11] SOUNDARARAJAN R., BOVIK A. C.: RRED indices: Reduced reference entropic differencing for image quality assessment. *IEEE Transactions on Image Processing 21*, 2 (2011), 517–526. 2

[SBC05] SHEIKH H. R., BOVIK A. C., CORMACK L. K.: No-reference quality assessment using natural scene statistics: JPEG2000. *IEEE Transactions on Image Processing 14*, 11 (2005), 1918–1927. 2

[SHS*04] SEETZEN H., HEIDRICH W., STUERZLINGER W., WARD G., WHITEHEAD L., TRENTACOSTE M., GHOSH A., VOROZCOVS A.: High dynamic range display systems. *ACM Transactions on Graphics 23*, 3 (2004), 760–768. 2

[Sim05] SIMONCELLI E. P.: Statistical modeling of photographic images. In *Handbook of Image and Video Processing*, Bovik A., (Ed.). Academic Press, May 2005, ch. 4.7, pp. 431–441. 2nd edition. 2

[SKMS06] SMITH K., KRAWCZYK G., MYSZKOWSKI K., SEIDEL H.-P.: Beyond tone mapping: Enhanced depiction of tone mapped hdr images. *Comput. Graph. Forum 25*, 3 (2006), 427–438. 2

[SKY*12] SEN P., KALANTARI N. K., YAESOUBI M., DARABI S., GOLDMAN D. B., SHECHTMAN E.: Robust patch-based hdr reconstruction of dynamic scenes. *ACM Transactions on Graphics 31*, 6 (2012), 203. 2, 10

[SL12] SILK S., LANG J.: Fast high dynamic range image deghosting for arbitrary scene motion. In *Proceedings of Graphics Interface 2012* (2012), Canadian Information Processing Society, pp. 85–92. 5, 7, 9, 10

[SPLC13] SUNG H.-S., PARK R.-H., LEE D.-K., CHANG S.: Feature based ghost removal in high dynamic range imaging. *International Journal of Computer Graphics & Animation 3*, 4 (2013). 2

[SSM12] SRIKANTHA A., SIDIBÉ D., MÉRIAUDEAU F.: An SVD-based approach for ghost detection and removal in high dynamic range images. In *21st International Conference on Pattern Recognition (ICPR)* (2012), IEEE, pp. 380–383. 2, 10, 11

[ST04] SAND P., TELLER S.: Video matching. In *ACM Transactions on Graphics (TOG)* (2004), vol. 23, ACM, pp. 592–599. 2

[TAEE15] TURSUN O. T., AKYÜZ A. O., ERDEM A., ERDEM E.: The State of the Art in HDR Deghosting: A Survey and Evaluation. *Computer Graphics Forum* (2015). 2, 8, 9

[TH94] TEO P., HEEGER D.: Perceptual image distortion. In *IEEE International Conference on Image Processing* (1994), IEEE. 2

[TJK11] TANG H., JOSHI N., KAPOOR A.: Learning a blind measure of perceptual image quality. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE* (2011). 2

[War03] WARD G.: Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. *Journal of Graphics Tools 8*, 2 (2003), 17–30. 3

[WB06] WANG Z., BOVIK A. C.: *Modern Image Quality Assessment*. Morgan and Claypool Publishers, 2006. 2

[WBSS04] WANG Z., BOVIK A. C., SHEIKH H. R., SIMONCELLI E. P.: Image quality assessment: From error visibility to structural similarity. *IEEE Transactons on Image Processing 13*, 4 (2004), 600–612. 2

[WR05] WU H. R., RAO K. R.: *Digital Video Image Quality and Perceptual Coding (Signal Processing and Communications)*. CRC Press, Inc., Boca Raton, FL, USA, 2005. 2

[YKKD12] YE P., KUMAR J., KANG L., DOERMANN D.: Unsupervised feature learning framework for no-reference image quality assessment. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE* (2012). 2

[ZC10] ZHANG W., CHAM W.-K.: Gradient-directed composition of multi-exposure images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (june 2010), pp. 530 –536. 2

[ZZMZ11] ZHANG L., ZHANG D., MOU X., ZHANG D.: FSIM:a feature similarity index for image quality assessment. *IEEE Transactions on Image Processing 20*, 8 (2011), 2378–2386. 2