The State of the Art in HDR Deghosting: A Survey and Evaluation

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

Abstract

Obtaining a high quality high dynamic range (HDR) image in the presence of camera and object movement has been a long-standing challenge. Many methods, known as HDR deghosting algorithms, have been developed over the past ten years to undertake this challenge. Each of these algorithms approaches the deghosting problem from a different perspective, providing solutions with different degrees of complexity, solutions that range from rudimentary heuristics to advanced computer vision techniques. The proposed solutions generally differ in two ways: (1) how to detect ghost regions and (2) what to do to eliminate ghosts. Some algorithms choose to completely discard moving objects giving rise to HDR images which only contain the static regions. Some other algorithms try to find the best image to use for each dynamic region. Yet others try to register moving objects from different images in the spirit of maximizing dynamic range in dynamic regions. Furthermore, each algorithm may introduce different types of artifacts as they aim to eliminate ghosts. These artifacts may come in the form of noise, broken objects, under- and over-exposed regions, and residual ghosting. Given the high volume of studies conducted in this field over the recent years, a comprehensive survey of the state of the art is required. Thus, the first goal of this paper is to provide this survey. Secondly, the large number of algorithms brings about the need to classify them. Thus the second goal of this paper is to propose a taxonomy of deghosting algorithms which can be used to group existing and future algorithms into meaningful classes. Thirdly, the existence of a large number of algorithms brings about the need to evaluate their effectiveness, as each new algorithm claims to outperform its precedents. Therefore, the last goal of this paper is to share the results of a subjective experiment which aims to evaluate various state-of-the-art deghosting algorithms.

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

1. Introduction

The real world encompasses a wide range of luminance values that exceeds the capabilities of most image capture devices. However, in general it is desirable to capture, store, process, and display this wide range of luminance values. The field of HDR imaging is primarily developed to address this problem, that is to bridge the gap between what is available in the real-world in terms of light levels and what we can do to represent it using digital equipment [RWPD10].

The first stage of the HDR imaging pipeline is *acquisition*. There have been many studies in HDR image and video acquisition, which can be grouped under three categories. The first category consists of the methods that use specialized hardware to directly capture HDR data. The second category consists of the techniques based on reconstructing an HDR image from a set of low dynamic range (LDR) images of the scene with different exposure settings, techniques that are collectively called as multiple exposure methods. The third category consists of the techniques which aim to expand the dynamic range of a normally LDR image – be it through pseudo-multi-exposure or inverse tone mapping [BADC11].

In general, the techniques in the first and third categories produce inherently ghost-free HDR images as they operate on data captured at a single time instance. The techniques in the second category, however, must deal with moving objects as the image capture process takes a longer time due to necessity of capturing multiple exposures. This is due to the fact that the ensuing HDR image reconstruction process sim-

(a) A moving light source with high noise

(b) Non-deformable body motion with large displacements

(c) Deformable body motion

(d) Deformable body motion with occlusions
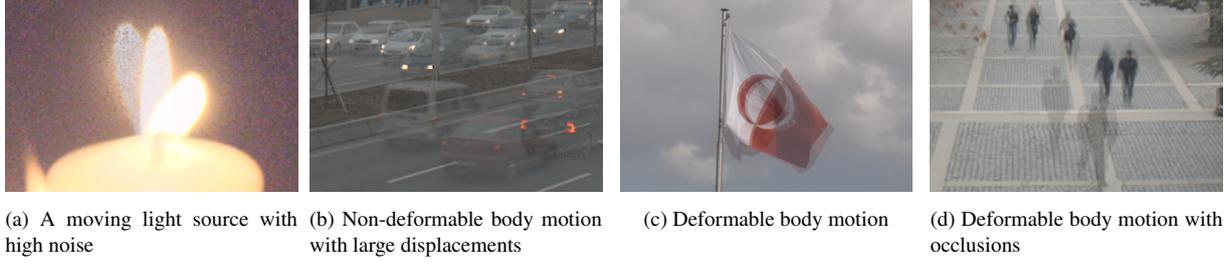
Figure 1: Different types of ghost artifacts.

Table 1: Notation used in this survey

$L_1, L_2, ..., L_N$	Input LDR images
L_{ref}	Reference LDR image
$L_n(p)$	Pixel intensity at position (p) in image L_n
H	Output HDR image
$f(\cdot)$	Camera Response Function
$g_{nm}(\cdot)$	Intensity Mapping or Color Transfer Function from exposure n to m
Δt_n	Exposure time of the input image L_n
$\alpha(L_n(p))$	Weight of pixel $L_n(p)$
$E_1, E_2, ..., E_N$	Input LDR image in the radiance domain (Eq. 2)

ply computes a weighted average of all exposures, resulting in different objects being blended together in case of object movement. The artifacts that occur as a result of such blending are collectively termed as *ghosts* or *ghosting artifacts* (see Figure 1).

We can formalize this notion as follows (see Table 1 for the terminology used in this survey). Let $L(p)$ represent an LDR image pixel p which is obtained when the corresponding sensor location is exposed to an irradiance $E(p)$ for Δt units of time (see Table 1):

$$L(p) = f(E(p) \cdot \Delta t), \qquad (1)$$

where f represents the camera response function (CRF) which depends on several factors such as the white balance and gamma correction setting, analog-to-digital conversion parameters, physical characteristics of the sensor, camera manufacturer preferences, etc. If the function f is known, it is possible to recover the correct sensor irradiance from the image pixel intensity using the following relation:

$$E(p) = \frac{f^{-1}(L(p))}{\Delta t}. \qquad (2)$$

Most of the time, f is not known but can be recovered using various techniques [BK93, MP94, DM97, RBS99, MN99, GN03, OAG13]. Alternatively, the images can be captured in RAW formats which are typically linear (thus $f(x) = mx$ for an easily recoverable slope value, m).

Once f is recovered, the HDR value $H(p)$ can be computed as:

$$H(p) = \frac{\sum\limits_{n=1}^{N} \alpha(L_n(p)) \frac{f^{-1}(L_n(p))}{\Delta t_n}}{\sum\limits_{n=1}^{N} \alpha(L_n(p))}, \qquad (3)$$

where α is a weighting function which depends on the pixel intensity level. Although one can use a simple triangular weighting function that gives high weights to the center of the intensity range while penalizing the extremes as proposed by Debevec and Malik [DM97], recent research has shown that other parameters such as the camera noise must be taken into account to determine an optimal weighting function [GAW*10].

The critical assumption of Equation 3 is that all input images $L_1, ..., L_N$ measure the same scene radiance value for each pixel position p:

$$\frac{f^{-1}(L_n(p))}{\Delta t_n} = \frac{f^{-1}(L_m(p))}{\Delta t_m} \quad \forall n, m, p. \qquad (4)$$

If this assumption, known as *reciprocity*, does not hold, $H(p)$ will be equal to the weighted sum of different sensor irradiance values, resulting in semi-transparent object appearances known as ghosts. The reciprocity assumption may break down for saturated pixels – a problem that is to be dealt with by using a good α function.

The requirement of a pixel measuring the same irradiance in all input exposures necessitates that the camera and the scene remain static throughout the capture process. Because this requirement would severely limit the range of scenes that can be captured using the multiple exposures technique, various solutions have been proposed for dealing with both camera and object movement. It should noted however that, of these two problems, the object movement is much more severe as the former can be avoided by using a tripod or registering the individual exposures prior to HDR image reconstruction. Although this survey discusses both types of methods, the emphasis is especially placed on deghosting algorithms that deal with dynamic objects.

2. Related Work

Despite the large number of HDR deghosting methods that have been proposed, there exists only a few studies that aim to survey, classify, and compare them. In this section, we briefly review these studies.

Srikantha and Sidibé [SS12] provided the first survey and classification of HDR deghosting methods. In their study, they classified 17 algorithms according to how they detect motion regions and remove ghosts. In addition, the ghost detection accuracy is evaluated quantitatively in terms of sensitivity and specificity by comparing the ghost detection bitmaps with the ground truth. Although this study is important as being the first review of deghosting methods, it does not cover a large number of methods and is now outdated since it misses the surge of activity that took place in the field after 2011.

Karaduzovic-Hadziabdic et al. [KHTM13] proposed a methodology for evaluating deghosting algorithms and compared the methods of Sen et al. [SKY*12] and Zimmer et al. [ZBW11] together with two commercial products, namely Photomatix and Photoshop, in a psychophysical experiment with 30 subjects and 9 scenes. It was found that Sen et al.'s algorithm has the fewest artifacts. In their more recent work [HTM14], the authors extended their comparison to include the algorithm of Hu et al. [HGPS13] and performed an expert evaluation.

In another study, Tursun et al. [TAEE14] compared the methods of Sen et al. and Hu et al. on a new dataset using both subjective and objective metrics. The selected objective metric was Liu et al.'s [LWC*13] no-reference metric for evaluating the quality of motion deblurring. The authors hypothesized that due to the similarity of deblurring and deghosting artifacts, the metric proposed for the former may be employed for the latter. However, the authors did not correlate the subjective and objective results.

The primary goal of the current study is to conduct a much more extensive and up-to-date survey and classification of the state of the art in the field of HDR deghosting. To this end, this survey includes approximately 50 HDR deghosting methods grouped into a taxonomy based on the approaches they follow. Secondly, due to the increasing number of deghosting algorithms that are being proposed each year, there is a growing need to evaluate these methods systematically. To this end, there is a need to have a reliable benchmark dataset for HDR image deghosting as it exists in other fields of computer vision such as optical-flow [BSL*11], image retrieval [SB05], and image retargeting [RGSS10]. Thus, the second goal of this paper is to present a new benchmark dataset and the results of a subjective experiment aimed to compare several deghosting algorithms. Finally, this paper makes a brief investigation of whether two somewhat relevant quantitative metrics can be used to objectively evaluate deghosting quality. In summary, the current work makes the following contributions:

- A taxonomy and review of a much more extensive set of deghosting studies available in the recent literature,
- A new benchmark dataset which can be used to test HDR image deghosting methods,
- A subjective evaluation framework for evaluating future deghosting algorithms,
- Comparison of 6 (relatively recently proposed) deghosting algorithms both in a subjective experiment and using two quantitative metrics.

The rest of this paper is organized as follows. In Section 3 we give a taxonomy of existing HDR deghosting methods and analyze the advantages and weaknesses of each class of algorithms. In Section 4, we first introduce the benchmark dataset and then provide the details of our experimental design and data analysis. In Section 5, we present the results of the experiment and interpret the obtained rankings based on the algorithms' outputs. We also discuss the correlation of two quantitative metrics with the experimental results. In Section 6, we conclude the paper with a summary of the findings and a discussion of future research directions.

3. Taxonomy of HDR Deghosting Methods

In order to organize and highlight the similarities and differences between existing deghosting algorithms, here we present a taxonomy of approximately 50 methods based on how they approach the deghosting problem.

Global exposure registration methods aim to align individual exposures globally.

Moving object removal methods aim to remove the motion by estimating a static background.

Moving object selection methods detect the inconsistencies in the input pixel intensities which are affected by motion and remove the ghosting artifacts by either locally using a *single source* image or combining a set of *multiple sources* which are consistent.

Moving object registration methods focus on recovering or reconstructing the ghost pixels by searching for the best matching region in other exposures or in the affected image. The matching regions are used to transfer information to the problematic region. These algorithms may find *pixel* or *patch* based dense correspondences.

Video deghosting methods are tailored to remove the potential ghosting artifacts in HDR videos. In this regard, they make use of the temporal information of videos during processing.

The detailed taxonomy is given in Figure 2 with accompanying references. It should be noted that although this taxonomy is valid for most cases, there exist some hybrid approaches which are difficult to classify as belonging to a single class. Such algorithms are classified based on their most dominant characteristics.

HDR Motion Compensation Methods

Global Exposure Registration	Moving Object Removal	Moving Object Selection			Moving Object Registration		Video Deghosting
		Single Source	Multi Source		Optical-flow based	Patch based	
[MMF02]	[KAR06]	[KHC*06]	[GGC*09]	[MPC11]	[Bog00]	[MG07]	[KUWS03]
[Can03]	[PH08]	[Gro06]	[MPC09]	[MTCL12]	[HG11]	[POK*11]	[ST04]
[War03]	[GSL08]	[JLW08]	[RKC09]	[SSM12]	[ZBW11]	[ZLZR11]	[MG10]
[CH06]	[SPS*09]	[LC09]	[RC10]	[ZC12b]	[FBPC12]	[OMLV12]	[CCCV11]
[EUS06]	[SL12]-PWD	[RWPD10]	[LRZ*10]	[OLK13]	[JO12]	[HGP12]	[CCV11]
[GG07]	[ZC10]	[PK10]	[WXRL10]	[SPLC13]	[HDW14]	[SKY*12]	[KSB*13]
[TM07]	[ZC12a]	[LPC11]	[HLL*10]	[GKTT13]		[HGPS13]	
[RMVS07]		[SL12]-FM	[ALKC11]	[WT13]		[ZLZ*12]	
[ILP11]			[RC11]	[LLM14]		[ZLZ*13]	
[Aky11]							
[Yao11]							

Figure 2: Taxonomy of HDR motion compensation methods

3.1. Global Exposure Registration

Although there are a few exceptions, almost all ghost removal methods assume that the input exposures are pre-registered either by capturing images by using a static camera or by using one of the methods discussed here. The objective of *global registration* algorithms is to compensate for the effects of the camera motion by estimating the parameters of transformations which will be applied to each one of the input images. These methods do not focus on the object displacements or assume that the scene is static. The types of the transformations and the methods of parameter estimation are the discriminative properties of the algorithms in this class.

In their early work, Mann et al. [MMF02] use global homography to align differently exposed frames of a panning video with unknown relative exposure settings to produce an image mosaic. They propose three methods based on comparametric equations for the simultaneous estimation of the CRF and image registration. Each frame is modeled as a function of quantity of light falling on the sensor where the functions are in the form of projective coordinate transformations with unknown parameters. The first method is based on minimizing the sum of squared errors between the irradiance values with smoothness and monotonicity constraints on the CRF. The second method estimates the CRF as a two-parameter closed-form function. While being simple, this method provides a fit with lower accuracy. The third method applies spline interpolation to the data points produced by a procedure called "log-unrolling". The output image obtained using these methods is as good as input images where there is no overlap between the images. In the regions where there is overlap, the output image is better than the individual input images since it combines the additional information from multiple images.

Candocia [Can03] introduces another method based on comparametric equations similar to Mann et al.'s method

[MMF02], where spatial and tonal registration of the input images are performed simultaneously. The parameters minimizing the variance of the pixel values at all spatial image coordinates are found using the Levenberg-Marquardt algorithm. Since this optimization is nonlinear with multiple local minima, the initial parameter set is initialized by pairwise registering the input images using pyramidal decomposition of the images. While being successful at correcting registration errors, the author states that the method may suffer from increasing running times due to the computation of the Hessian matrix at each iteration of the optimization.

Ward [War03] presents a fast translational registration of different exposures by aligning bitmaps obtained from the input images. A Median Threshold Bitmap (MTB) M_i of image L_i is defined as:

$$M_i(p) = \begin{cases} 1, \text{if } L_i(p) < \mu_i \\ 0, \text{otherwise} \end{cases} \quad (5)$$

where μ_i is the median value of the pixel intensities in image L_i. Ward's approach is based on the observation that M_i is robust to the changes in the exposure settings, due to the monotonicity property of camera response functions. In the study, each M_i, M_j pair is aligned using a multi-scale pyramid structure, starting from the lowest resolution. Later, the translations found using the pyramid are applied to the input images to obtain a co-registered set of input images. The method of Ward does not require the CRF estimation and the computational cost is very low since it is possible to process multiple bits of the bitmaps in one clock cycle of the CPU.

Cerman and Hlaváč [CH06] introduce an HDRI method to estimate the unknown exposure time from a set of RAW images. As a part of their work, they present a method to register the input images in order to eliminate the camera motion present in the hand-held acquisition settings. Prior to

the image registration, the input exposures are normalized using the estimated CRF. Later, the amount of image shift is estimated using the correlation in the Fourier domain. The scope of this initial estimation is limited to translational camera motion only. The estimates are used to initialize the local optimization of the sum of squared differences between the input images. The optimization phase includes both translational and rotational motion in subpixel accuracy. Instead of choosing a reference image, the registration is applied on consecutive image pairs.

Eden et al. [EUS06] present a method for the mosaicing of images with large exposure difference and scene motion to create an HDR panorama. As the first step, the input images are aligned using feature-based registration method of Brown and Lowe [BL03]. Next, the input images are mapped to the radiance domain. In this procedure, a precalibrated camera with a known CRF is used. The output panorama is constructed in two phases. In the first phase, a reference panorama is constructed based on Agarwala et al. [ADA*04]. Although the created reference panorama covers the full angular extent with smooth transitions between the input images, it does not use the full dynamic range. In the second phase, the full dynamic range is introduced to the reference panorama using a max-flow graph cut which encourages large SNR while preserving the smooth transitions between the images. Since each pixel of the output is constructed using only one of the input images, potential ghosting artifacts are eliminated.

Gevrekci and Gunturk [GG07] propose a novel contrast-invariant feature transform (CIFT) algorithm which does not require a photometric registration as a preprocessing step of the spatial registration. Based on the assumption that the Fourier components are in phase at the corners, the algorithm detects the corners by applying a local contrast stretching operation to each pixel of an input image and using the Phase Congruency (PC) function [Kov99]. Next, the input images are spatially registered by feature matching using RANSAC.

Tomaszewska and Mantiuk [TM07] aim to correct misalignments due to the camera motion by estimating a general planar homography using SIFT [Low04] features and RANSAC [FB81]. First, the proposed algorithm extracts SIFT keypoints in the input images. Second, the correspondences between the keypoints are established and the number of correspondences is decreased to four pairs using the RANSAC algorithm. The RANSAC algorithm selects the keypoints which are compatible with a homography and which are present in all of the input images. Since the transformation provided by the homography has subpixel accuracy, the output pixel values are calculated using bilinear interpolation. In addition, the keypoint search procedure is performed in the contrast domain to provide robustness against the changes in the exposure and using a multi-scale difference-of-Gaussian pyramid to improve the keypoint detection. In order to increase the accuracy of the registration

process, SIFT algorithm is modified to choose an automatic threshold value instead of a fixed one.

Rad et al. [RMVS07] start with estimating the CRF, f, based on Debevec et al. [DM97]. Then f^{-1} is applied to the input images to transfer the images to the radiance domain. Alignment of the input images is performed in the frequency domain, using Fourier transform and ignoring the pixels closer to the limits of the pixel intensity range. Different from Cerman and Hlaváč [CH06], this method estimates both the translation and the rotation parameters in the frequency domain only. Later, these parameters are used to create a super-resolution HDR image by interpolating in a higher-resolution image grid in the HDR reconstruction phase.

Im et al. [IJLP11] propose an algorithm based on the elastic registration (ER) method of Periaswamy et al. [PF03]. In their study, Im et al. model the motion between the consecutive image pairs using affine transformations. The transformation parameters are estimated by minimizing the sum of squared differences between the pixel intensities of transformed image pairs. Their other work [ILP11] propose an improvement to the ER by selecting the best target image for the registration. The target image is selected by averaging the hue channel of the input images. Later, the image which has the smallest mean of squared error with this average hue map is chosen as the target image. This operation improves the estimation accuracy of the ER by avoiding the selection of over- and under-exposed images as targets.

Akyüz [Aky11] eliminates the misregistrations due to the translational camera motion only. The proposed approach is based on the observation that unless they are not under- or over-exposed, the relation between the intensities of neighboring pixels in a set of aligned images (e.g. a pixel having smaller intensity than its bottom neighbor and larger intensity than its right neighbor) should be insensitive the exposure changes. In other words, pixel order relations should be preserved. For each input image, a correlation map is created which maps such relations to ordinal values. Later, input images are aligned pairwise by minimizing the Hamming distance between the correlation maps of the input images. Rotation and other more complex camera motion patterns are not addressed by this method.

Yao [Yao11] provides another method based on the use of phase congruency (PC) images. Different from Gevrekci and Gunturk [GG07], the proposed method registers PC images in the frequency domain using the phase cross-correlation technique, rather than using them to identify the keypoints for a registration in the spatial domain. In addition to the translational registration, the rotational registration is also performed using log-polar coordinates, in which rotational motions are represented with translational transformations in the coordinates. In order to detect the subpixel shifts, evolutionary programming is used to find the optimal transformation parameters.

3.2. Moving Object Removal

The main objective of the algorithms in this class is to re-move all of the moving objects in the scene by estimating the static background. One important assumption of the most of these algorithms is that for each pixel, the majority of the input exposures capture the static part of the scene. Due to this assumption, insufficient number of input exposures, dynamic backgrounds, and deformable-body motions with overlapping regions between exposures have a negative ef-fect on the deghosting quality for this class of approaches.

Khan et al. [KAR06] propose an iterative method for the removal of moving objects in the output HDR image. The method is based on the calculation of the probability $P(E_n(p)|F)$ that a pixel $E_n(p)$ belongs to the background class F. The authors use the kernel density estimator:

$$P(E_n(p)|F) = \sum_{i=1}^{j} w_i \cdot K_{\mathbf{H}}(E_n(p) - \mathbf{y}_i) / \sum_{i=1}^{j} w_i, \quad (6)$$

where j is the number of pixel neighbors and $K_{\mathbf{H}}$ is the multi-variate kernel function:

$$K_{\mathbf{H}}(\mathbf{x}) = |\mathbf{H}|^{-\frac{1}{2}} (2\pi)^{-\frac{d}{2}} \exp(-\frac{1}{2}\mathbf{x}^T \mathbf{H}^{-1}\mathbf{x}), \quad (7)$$

where $\mathbf{H}$ is the bandwidth matrix. In the calculation of Equa-tion 6, $\mathbf{y}_i$ belongs to 3×3 neighborhood of the pixel p in *all* exposures. w is initially set to a hat function then both w_p and $P(E_n(p)|F)$ are updated iteratively until convergence:

$$w_{p,t+1} = w_{p,0} \cdot w_{p,t}, \quad (8)$$

where t is the iteration number. At the end of each iteration, $w_{p,t}$ takes the new value of $P(E_n(p)|F)$. After the conver-gence, the probabilities $P(E_n(p)|F)$ are used instead of the weighting function α in Equation 3 to construct the HDR image.

Pedone and Heikkilä [PH08] improve the method of Khan et al. [KAR06] by estimating the parameters of the band-width matrix $\mathbf{H}$. Instead of setting it to the identity matrix, the i-th diagonal entry of $\mathbf{H}$ is calculated as:

$$\tilde{h}_i = n^{-\frac{1}{d+4}} \tilde{\sigma}_i, \quad (9)$$

where $\tilde{\sigma}$ is the weighted standard deviation of the ith di-mension. In addition, due to similarity of some objects with the background, the authors state that using Khan et al.'s al-gorithm [KAR06], it is possible to observe ghosting in the output. To improve the deghosting quality, they propose to apply morphological operations to the bitmaps obtained by thresholding the weight maps before merging the exposures.

Granados et al. [GSL08] provide an energy-minimization based background estimation method whose application to HDRI is also given. For each pixel p, the proposed method assigns the source image a label $\mathcal{L}_p \in \{1,...,N\}$ which min-imizes:

$$\mathcal{E}(\mathcal{L}) = \sum_p D_p(\mathcal{L}_p) + \sum_{(p,q) \in \mathcal{N}} V_{p,q}(\mathcal{L}_p, \mathcal{L}_q)$$
$$+ \sum_{(p,q) \in \mathcal{N}} H_{p,q}(\mathcal{L}_p, \mathcal{L}_q), \quad (10)$$

where $D_p(\mathcal{L}_p)$ is the data term, $V_{p,q}(\mathcal{L}_p, \mathcal{L}_q)$ is the smooth-ness term, and $H_{p,q}(\mathcal{L}_p, \mathcal{L}_q)$ is the hard constraint. The data term measures how well p satisfies the estimated density function and the approximated motion boundaries. The smoothness term $V_{p,q}(\mathcal{L}_p, \mathcal{L}_q)$ penalizes intensity dif-ferences. The hard constraint $H_{p,q}(\mathcal{L}_p, \mathcal{L}_q)$ prevents half-included objects by allowing only previously observed la-beling transitions. The energy function in Equation 10 is minimized via graph cuts [BVZ01]. Resulting labeling de-termines the input source image for each pixel in HDR con-struction.

Sidibe et al. [SPS*09] detect ghost regions using the pixel order relation. If a pixel does not contain scene motion then the pixel intensity values must follow the same order as the exposure times, i.e. if $\Delta t_i > \Delta t_j$, then $L_i(p) \geq L_j(p)$. For each pixel in the ghost-regions, the LDR images are put into two sets D and S. S contains the exposures with no motion for current pixel and D contains the exposures with motion for the current pixel. Quasi Continuous Histograms (QCH) [CS07] are used to separate the input images into these two sets. The main approach of QCH is to calculate the mode of the pixel intensity values and use this informa-tion to identify the pixels with motion. During HDR image construction phase, only the images in the set S are used.

Silk and Lang [SL12] introduce a method which employs two different strategies depending on the type of the mo-tion. The algorithm starts with performing a change detec-tion which consists of applying a fixed threshold on the ab-solute difference of irradiance values in each color channel. This initial motion mask does not respect the object bound-aries. In order to refine the initial mask according to the object boundaries, the images are first over-segmented us-ing SLIC superpixels [ASS*10] and then the superpixels are categorized into motion and non-motion regions according to the number of inconsistent pixels marked by the initial change detection. The super-pixels with motion are assigned smaller weights in the HDR reconstruction. This operation is called pairwise down-weighting (PWD). The algorithm has a second type of output in the presence of fluid motion (FM), which is described in Section 3.3.1.

Zhang and Cham [ZC10,ZC12a] propose an exposure fu-sion algorithm of static and dynamic scenes, where the pixel weights are determined using gradient-domain based quality measures instead of absolute pixel intensities. The per-pixel weight map of each image is calculated as a multiplication of *visibility* and *consistency* scores. The visibility score assigns larger weights to the pixels with larger gradient magnitudes. On the other hand, consistency score assigns larger weight

to a pixel if its gradient direction is consistent with the col-located pixels in other exposures.

3.3. Moving Object Selection

The algorithms that fall into this class are characterized by the approaches they employ to detect the presence of motion, e.g. variance-based, pixel-value prediction, thresholding, pixel-order relation, etc. Different from the *moving object registration* methods, the *moving object selection* does not compute correspondences among the input LDR images to recover the pixel intensities using all exposures while compensating for motion. Instead, they select one (single source) or multiple (multi source) input images for each dynamic region. Hence, while computationally efficient, moving object selection algorithms, in particular single source ones, have a drawback that the resulting image may not be HDR in dynamic regions.

The major difference of the *moving object selection* methods from the *moving object removal* methods is that the former select one or more source images for the regions affected by motion. As a result, the output HDR contains the moving objects which appear in the selected input image(s). On the contrary, *moving object removal* methods do not select a particular reference image. They perform a consistency check for each pixel across exposures. This results in complete removal of dynamic objects if they do not stay stationary in the majority of the input exposures.

3.3.1. Single Source

Single source methods use a single input image for each dynamic region. Some simpler methods use the same input image for all dynamic regions whereas more sophisticated ones may choose a different input image for different dynamic regions based on a well-exposedness criteria.

Kao et al. [KHC*06] works on two RAW LDR images with ± 2 EV difference. First, a block-based global alignment is performed which removes the effects of camera motion. Due to the ± 2 EV difference the following relation between the pixel intensity value of two input images L_1 and L_2 is expected:

$$L_2(p)/L_1(p) = 4, \quad \Delta t_2 = 4\Delta t_1. \tag{11}$$

If a pixel is not consistent with this relation (excluding saturated pixels), it is marked as a potential ghost. Next, the exposure normalized version of the low exposure image is calculated as $\tilde{L}_1(p) = 4 \cdot L_1(p)$, and the output HDR image H is then obtained by fusing the input images as follows:

$$H(p) = \begin{cases} \tilde{L}_1(p), & \text{if } p \text{ is marked or } L_2(p) \text{ is ill-exposed} \\ L_2(p), & \text{otherwise.} \end{cases} \tag{12}$$

Grosch [Gro06] provides an extension to Median Threshold Bitmaps (MTB) [War03] by including rotational alignment

where MTB alignment is performed on the graphics hardware to accelerate the computations. This first stage of utilizing the MTBs prevents the potential artifacts caused by the camera movement. The next step includes a CRF estimation which is based on the histogram-based method of Grossberg and Nayar [GN03] and used to predict pixel intensity values of each image in consecutive image pairs L_i and L_j. A pixel p is marked as ghost if the following relation does not hold:

$$\left| f\left(\frac{\Delta t_j}{\Delta t_i} f^{-1}(L_i(p)) \right) - L_j(p) \right| < \varepsilon \tag{13}$$

The pixels of the ghost regions are not used in the HDR image construction. In order to minimize the loss in the dynamic range and noise, the pixel intensities in these regions are predicted using the estimated CRF and the intensities of the co-located pixels from the source image with the lowest number of poorly-exposed pixels in the motion region.

The algorithm of Jacobs et al. [JLW08] consists of two steps. First, similar to [Gro06], a global image alignment is performed based using MTBs [War03]. Next, the ghosting artifacts caused by moving objects are eliminated by making use of the so-called *variance images* (VI) and local-entropy based *uncertainty images* (UI). The VI is created by calculating the per-pixel intensity variance over all exposures in the radiance domain, excluding the saturated exposures. To have better region boundaries, the VI is thresholded with a fixed value and morphological operations are applied to the resulting bitmap. Since the calculation of the VI is performed in the radiance domain, inaccurate CRF estimation may result in unreliable variance values. Therefore, the authors use the *uncertainty images* (UI) as a supplementary movement detection source since the UI does not require the CRF. The use of UI is based on the assumption that local contrast sources such as edges correspond to object boundaries and the entropy around these regions should be similar if the region is not affected by the scene motion. The final UI is created by taking a weighted difference of UI corresponding to each input image, and then by applying thresholding and morphological operations to obtain motion region clusters. In the HDRI generation phase, the source image used in the movement regions is chosen as the input image with the least amount of saturation and the longest exposure time.

Lin and Chang [LC09] propose a method to eliminate ghosting artifacts caused by stereo mismatches in stereoscopic HDR. The input is a pair of images captured with different exposure time settings using two cameras. The disparity map between the input images are found using the SIFT matching scheme after normalizing input images with the estimated CRF. If the absolute difference between the corresponding pixels of the normalized images are larger than a threshold in a region, it is identified as a ghost region caused by a stereo mismatch.

Reinhard et al. [RWPD10] calculate the weighted variance at each pixel location and selects the regions where the variance is above a threshold as a motion region. The method

works well when the moving object is significantly different from the background in terms of the contrast and color similarity.

Pece and Kautz [PK10] propose a motion-region detection approach called Bitmap Movement Detection (BMD) which is also based on MTBs [War03]. They start by extracting MTB M_i of each input exposure L_i. Then any pixel p, for which the following is true, is marked in the motion map:

$$\sum_{i=1}^{N} M_i(p) \notin \{0, N\} \qquad (14)$$

To remove the effects of the noise, the obtained motion map is refined using morphological operations such as erosion and dilation. The pixels in the motion map are clustered according to their connectivity using the connected component labeling algorithm of Haralick and Shapiro [SH92], and finally, the detected motion regions are filled from the best-exposed input image for each individual motion region.

Lee et al. [LPC11] propose a histogram-based deghosting method which improves the studies of Min et al. [MPC09, MPC11] by detecting ghost regions as a difference in the ranks of pixels according to their intensities. In order to be able to scale the computational load, the pixel ranks are normalized to B bits using:

$$\tilde{r}_n(p) = \texttt{round}\left(\frac{r_n(p) - 1}{R_n - 1} \times 2^B\right), \ 0 \le \tilde{r}_n(p) \le 2^B - 1, \qquad (15)$$

where $r_n(p)$ is the rank of the pixel p in L_n and $\tilde{r}_n(p)$ is the normalized rank. Larger B gives smaller quantization error with higher computational load. If the absolute difference of normalized ranks between an image L_n and L_{ref} is larger than a user-defined threshold, the pixel is marked in a motion map. The rank-based motion maps are combined with the weighting function of Mertens et al. [MKVR07] so that only L_{ref} is used to produce output pixels in motion regions.

The algorithm of Silk and Lang [SL12] was introduced in Section 3.2 with its pairwise down-weighting (PWD) approach which is applicable when the minority of the input image stack is affected by the motion. However, in the presence of foliage, flags and fluids, some super-pixels may contain motion in each one of the input images. This motion type is called fluid motion (FM) and is not resolved correctly by the PWD. For such cases, the algorithm offers a second output which uses only the best exposure maximizing the sum of pixel weights in the region affected by the motion.

3.3.2. Multi Source

Multi source methods try to maximize dynamic range by using as many exposures as possible for each dynamic region. That is, different from *single source*, the input exposures which are consistent with a selected reference exposure contribute to the HDR image.

Gallo et al. [GGC*09] start by determining the reference image L_{ref}, which is selected either by the user or by minimizing the number of saturated pixels. Next, the ghost regions are found based on the reciprocity assumption:

$$\ln(L_n(i,j)) = \ln(L_m(i,j)) + \ln(e_{mn}), \qquad (16)$$

where e_{mn} is the relative exposure between L_m and L_n. This assumption states that there should be a linear relation if two pixels measure the same irradiance level. Any pixel violating this linear relation is considered as an inconsistent pixel containing scene motion. In order to increase the robustness, the method operates on rectangular image patches instead of the pixels. The inconsistent patches do not contribute to the HDR construction process. In order to avoid the artifacts between the patch boundaries, the HDR construction operation is performed in the image gradients domain [FLW02].

Min et al. [MPC09] extract multi-level threshold maps from each one of the input LDR images. A multi-level threshold map is a segmentation of the image into multiple regions according to the pixel intensity values where each region has the same number of pixels. Any difference between the threshold maps of input images and L_{ref}, which is selected as the mid-exposure is marked as a motion-region, and the pixels in the motion-regions are assigned smaller weights during HDR construction. While the proposed approach is simple and very fast, the presence of texture-less surfaces and the differences in the threshold maps due to other factors such as noise may result in false detections.

Raman et al. [RKC09] assumes that the first 5-10 horizontal lines of $\{L_1, ..., L_N\}$ do not contain motion, since motion is usually present in the ground plane of the scenes. These static regions are used to estimate the intensity mapping function (IMF). Similar to Gallo et al. [GGC*09], this approach checks the inconsistency between the input images and L_{ref} using rectangular patches. If a large number of pixels in a patch does not follow the IMF, the patch is marked as motion region in the source image and ignored in the exposure fusion operation [MKVR07].

Raman and Chaudhuri [RC10] improve the simple heuristic of Raman et al. [RKC09] by using a weighted variance measure based on Reinhard et al. [RWPD10] and Jacobs et al. [JLW08] to identify the static regions. The detected motion-free regions are used to estimate the IMF by fitting a polynomial curve with a degree of 4 to the observed pixel intensities in each pair of images. Then, each one of the input LDR images are over-segmented into super-pixels. A super-pixel is classified as a motion-region if the number pixels which do not follow the estimated IMF is above a certain level. The neighbor super-pixels with motion are merged and these regions are ignored, while the remaining static patches are merged using the exposure fusion technique of Mertens et al. [MKVR07].

Li et al. [LRZ*10] use a bidirectional pixel similarity measure between each LDR image and L_{ref} to identify the

regions affected by motion. For a particular pixel, if the pixel intensity measured in L_{ref} is more reliable (with an intensity value closer to 128), then the similarity is calculated by mapping the intensity of input image to L_{ref} using the IMF. Otherwise, it is calculated by mapping the intensity value of L_{ref} to the input image exposure. An adaptive threshold is applied to the similarity map in order to detect movement regions, which is a function of Δt and pixel intensities. The regions with motion are filled using the assumption that if for two pixels $L_{ref}(p) = L_{ref}(q)$, then for any source image $L_n(p) = L_n(q)$ must be satisfied.

Wu et al. [WXRL10] introduce a non-iterative ghost-free HDR imaging method without manual threshold tuning, which consists of alignment, movement detection, CRF estimation, and progressive image correction steps. The regions where the direction of the RGB vector remains fixed with respect to the exposure change are assumed to be free of motion and used for the CRF estimation. Later, this initial movement detection is refined using the pixel order relation:

$$\Delta t_i > \Delta t_j \implies L_i(p) \geq L_j(p), \qquad (17)$$

and the pixel error criterion, which is given by Equation 13 in Grosch [Gro06]. The refined movement detection mask is obtained by merging the binary movement masks of the color error criterion, pixel order relation, and the pixel error criterion. Then a progressive image correction is applied by starting from the reference exposure L_n and filling the motion regions with the predicted pixel values in the exposures L_{n-1} and L_{n+1}. The image correction is performed progressively for other exposures until all the input images are corrected. Formation of artifacts around object boundaries is prevented using the image inpainting technique of Olivera et al. [RC01].

Heo et al. [HLL*10] detect motion-regions using joint probability density functions (PDF) of pixel intensities from different exposures. After L_{ref} is selected, the global alignment of input images is performed to eliminate the effects of the camera motion. Next, the joint PDFs are estimated by applying Parzen windowing [Par62, R*56] to joint histograms between each pair:

$$\{\langle L_{ref}, L_n \rangle | L_n \in \{L_1, ..., L_N\} - \{L_{ref}\}\}. \qquad (18)$$

For each one of the nonreference images, a ghost bitmap is calculated by thresholding the joint PDF. Since the initial motion detection is noisy, it is refined with an energy minimization approach using graph cuts [BVZ01]. In the next phase, the CRF is estimated using the pixels in the static regions. The HDR reconstruction weights $\alpha(L_n(p))$ are based on bilateral filtering weights [TM98] and they are a function of pixel exposure, geometric distance and the color difference between $L_n(p)$ and $L_{ref}(p)$.

An et al. [ALKC11] propose another LDR exposure-fusion algorithm. Different from Raman et al. [RKC09] and

Raman and Chaudhuri [RC10], the motion detection operation is embedded into the exposure fusion equation with the following weighting formula:

$$\alpha(L_n(p)) = W_n(p) \cdot Z_n(p) \cdot O_n(p), \qquad (19)$$

where $W_n(p)$ is the weighting term used by Mertens et al. [MKVR07] which depends on contrast, saturation, and well-exposedness, $Z_n(p)$ is the zero-mean normalized cross correlation (ZNCC) factor between L_n and a previously selected L_{ref}, $O_n(p)$ is a binary map which is zero if $L_n(p) < L_m(p)$ when $\Delta t_n > \Delta t_m$.

In their more recent work, Raman and Chaudhuri [RC11] propose some improvements to their previous study [RC10] by replacing the exposure fusion approach of Mertens et al. [MKVR07] with a gradient domain solution. The fusion in the gradient domain is performed by placing larger weights to the pixels whose intensities are in the middle of the intensity range and which have higher local contrast. The effect of the noise on the local contrast is eliminated by smoothing the images.

In their more recent work, Min et al. [MPC11] improve their previous motion detection algorithm based on multi-level threshold maps [MPC09] and employ a noise reduction operation in the HDR reconstruction phase. As a preprocessing step, all input images are registered to L_{ref}. The false motion detection which occurs near the threshold values is eliminated by removing a pixel group in image L_n from the motion bitmap if the following two conditions hold:

1. The difference between the multi-level threshold maps of L_n and L_{ref} is low for that pixel group.
2. The pixel group is not a spatial neighbor of another pixel group whose multi-level threshold map is highly different from that of L_{ref}.

In the HDR reconstruction phase, Debevec and Malik's method [DM97] is modified by incorporating a down-weighting term for ghost regions and filtering for noise reduction using a structure adaptive anisotropic filter [YBFU96, MOW*07, MW06].

Moon et al. [MTCL12] handle the ghosting problem by introducing an additional term to the weighting formula of Mertens et al. [MKVR07], similar to An et al. [ALKC11]. First, a histogram matching operation is applied between each input image L_n and the reference image L_{ref}. Then, the ghost presence probability for each pixel is calculated as:

$$M_n(p) = \exp\left(-\frac{\left(L_{ref}(p) - \tilde{L}_n(p)\right)^2}{2c\sigma_{noise}^2}\right), \qquad (20)$$

where $\tilde{L}_n(p)$ is the pixel value obtained after applying the histogram matching operation to L_n, c is a user-set threshold and σ_{noise} is the image noise level. The value obtained as $M_n(p)$ is multiplied with the contrast, saturation, and well-exposedness terms of Mertens et al. [MKVR07] to obtain the enhanced output LDR.

Srikantha et al. [SSM12] propose a method which works on input images with linear CRF. Their work is based on the assumption that if the pixels from different exposures capture a static region of the scene, they must be linearly dependent since they are equal to the multiplication of sensor irradiance and exposure time. The pixels which do not follow the linearity and potentially cause ghosting are found using singular value decomposition (SVD) of a matrix containing pixel intensities from all exposures. This matrix is reconstructed using only the largest singular values, forcing the linearity between the corresponding pixel intensities of different exposures. The reconstructed pixel intensities are used to produce a ghost-free HDR image.

Zhang and Cham [ZC12b] improve their previous gradient-based deghosting method [ZC10, ZC12a] which assume that the majority of the pixels capture the static part of the scene for each motion region. Since this requirement is not satisfied for frequently changing scenes, here they introduce a consistency check with the pixels of the reference image instead of the majority of the exposures.

Oh et al. [OLK13] solve a rank minimization problem which simultaneously aligns the input images and detects moving objects together with ill-exposed regions. The proposed method works on input images with linear CRF. There are two assumptions used in this study. First, it is assumed that motion regions and under-/over-exposed pixels are sparse but cause large changes in the pixel intensities. Second, it is assumed that the camera motion is in the form of an homography transformation. With these assumptions each image L_n is represented as:

$$
\begin{aligned}
I_n \circ h &= f(k(R + S_n) \cdot \Delta t_i) \\
&= kR \cdot \Delta t_i + kS_n \cdot \Delta t_i \\
&= A_n + S_n,
\end{aligned}
\tag{21}
$$

where $I_n = L_n \circ h^{-1}$, $\circ$ is the element-wise mapping operator, h is the homography transformation, k is a scaling factor, f is the CRF, R is the sensor irradiance, and S_n is the sparse error term representing motion and the saturation. The matrix $\mathbf{A}$ and $\mathbf{S}$ are calculated by stacking the elements of each A_n and S_n column-wise, respectively. It is expected that $\mathbf{A}$ is a rank-1 matrix and all the artifacts are contained in the matrix $\mathbf{S}$ with $\mathbf{S} = 0$ in an artifact-free acquisition. The matrix of observed intensities $\mathbf{O} \circ \mathbf{h}$ is decomposed into a rank-1 matrix $\mathbf{A}$ and a sparse matrix $\mathbf{S}$. The result of the decomposition is used to recover the artifact-free observation of the scene.

Sung et al. [SPLC13] apply a local thresholding to the zero-mean normalized cross correlation (ZNCC) [TDSM07] maps, which is robust to the changes in illumination, to find the motion regions. After L_{ref} is selected, the translational and rotational alignment of the input images are performed using the SIFT-based approach of Tomaszewska and Mantiuk [TM07]. Next, the motion regions are detected with an adaptive local thresholding of ZNCC maps obtained from the luminance channels of the input images, excluding the

saturated pixels. In the HDR construction, the weights of the pixels in the motion regions are set to zero.

Granados et al. [GKTT13] introduce a Markov Random Field (MRF) based approach for ghost-free HDR imaging of dynamic scenes. In their study, if the camera motion is present, input images are aligned with a global homography using SURF key-points at first. Later for each pixel, consistent and inconsistent subset of input exposures are found by minimizing:

$$
\begin{aligned}
\mathcal{E}(F) = \sum_p \Big(\mathbf{1}_{\{Pr(p|F(p)) < \alpha\}} + \gamma \cdot V(F(p)) \Big) + \\
\beta \cdot \sum_{(p,q) \in \mathcal{N}} \mathbf{1}_{\{Pr(p|F(p,q)) < \alpha \vee Pr(q|F(p,q)) < \alpha\}},
\end{aligned}
\tag{22}
$$

where $\mathbf{1}$ is the indicator function, F is a mapping which assigns a set of input exposures as labels to each pixel, $\mathcal{N}$ is 4-neighborhood, and α, β, and γ are the user-set parameters. The first summation consists of consistency and noise potential terms while the second summation is the prior potential. The consistency and prior potentials penalize the inconsistent assignment of pixel labels by F. On the other hand, the noise potential penalizes the worsening of SNR in final HDR image due to trivial solutions to the energy function such as selecting only one image as a source and ignoring other input images. Using this proposed method, reference image selection and background estimation are not performed. The authors state that their method cannot recover the dynamic range of moving objects since moving objects are reconstructed from a single image. In addition, since there is not any semantic constraint in the HDR reconstruction, there may be inconsistencies such as object repetitions and half-included objects.

Wang and Tu [WT13] normalize the brightness level of all input images to the brightness level of the reference image L_{ref} in *Lab* colorspace. A ghost mask is obtained by thresholding the absolute difference of pixel intensities from each input image L_n and L_{ref}. The adaptive threshold $T_n(p)$ is given as:

$$
T_n(p) = |\Delta \bar{l}_n|^\beta + \left(\frac{L_n(p) - 50}{15} \right)^2,
\tag{23}
$$

where $\Delta \bar{l}_n$ is the average brightness difference and β is a user-selected tolerance factor. The ghost masks are refined using morphological operations and the complement of the ghost mask is used as the fourth term of the weight formula of Mertens et al. [MKVR07] to produce the output.

Lee et al. [LLM14] propose another rank minimization approach which is very similar to Oh et al. [OLK13]. The optimization function used in Lee et al. [LLM14] does not contain homography mappings but instead includes a separate variable as the ghost mask.

3.4. Moving Object Registration

Moving object registration methods focus on recovering or reconstructing the pixels affected by the movement by finding a local correspondence for the regions affected by motion. The main difference between the registration-based deghosting algorithms is their alignment strategy, such as feature matching (e.g. SIFT, Harris corner detector, etc.) or the alignment quality metric they use (e.g. Sum of Squared Differences, Cross-Correlation, etc.). Since the image registration task is a well-studied problem in other image processing domains, the set of algorithms in this class is very diverse and divided into subgroups. While *optical-flow based* approaches find a pixel-wise matching between the input images, *patch-based* methods use image patches and patch-based matching strategies to eliminate ghost regions. However, it should be noted that patch-based may also register individual pixels (similar to optical-flow) by computing the dense correspondence of overlapping patches around pixels.

3.4.1. Optical-flow based

The approaches in this group are mostly based on optical-flow estimation, which is a well-studied problem especially in stereo vision applications. In the HDR domain, optical-flow estimation must also take the exposure differences between the input images into account. The accuracy of the estimation is very critical for the quality of the outputs since any mismatch results in undesirable artifacts. In addition, use of optical-flow presents other challenges such as handling the occlusion, noise, or large displacements in the scene.

Bogoni [Bog00] introduces a pattern-selective fusion process which uses Laplacian pyramid [BA83] representation of the input images. This fusion process is very sensitive to the correct alignment of the input images. In order to prevent ghosting artifact due to motion, their method employs a two-phase alignment strategy. First, global affine transformation is performed to eliminate the effects of the camera motion. Second, optical-flow is estimated between the input images and L_{ref}. The use of Laplacian pyramid representation in both image fusion and optical-flow estimation decreases sensitivity to the changes in exposure.

Hossain and Gunturk [HG11] begin with estimating the intensity mapping function g_{nm} from the input image L_n to another input image L_m. Then the dense motion fields $\mathbf{u}_n$ are estimated using the optical-flow estimation algorithm of Zach et al. [ZPB07], minimizing the forward and backward flow residuals r_n and r_m:

$$r_n(p) = L_m(p - \mathbf{u}_n(p)) - g_{nm}(L_n(p)),$$
$$r_m(p) = L_n(p + \mathbf{u}_n(p)) - g_{mn}(L_m(p)). \qquad (24)$$

Starting with a static motion field $\mathbf{u}(p) = 0$, $\mathbf{u}$ and g are updated iteratively by minimizing the residuals until convergence. In order to obtain an estimate of g_{nm} which is robust to the effects of occlusion, each pixel p in each image L_n is

assigned an occlusion weight $w_n(p)$ with the following sigmoid function:

$$w_n(p) = 0.5 - \tan^{-1}((|r_n(p)| - \mu)/\pi\sigma), \qquad (25)$$

where μ and σ are the parameters controlling the shape of the function. The weights $w_n(p)$ measure the likelihood of the visibility of pixel p in the other image L_m. The intensity mapping functions are estimated on the weighted histograms using $w_n(p)$.

Zimmer et al. [ZBW11] present an energy-based method for estimating the optical-flow. This approach is claimed to be robust in the presence of noise and occlusion. One of the images in the input LDR set is selected as reference and dense displacement fields $\mathbf{u}_n$ between L_{ref} and each input image L_n is estimated by minimizing:

$$\mathcal{E}(\mathbf{u}_n) = \sum_p [D(\mathbf{u}_n) + \gamma S(\nabla \mathbf{u}_n)], \qquad (26)$$

where γ is the weighting coefficient, D is the data term measuring the quality of the alignment in the gradient domain and S is a spatial smoothness term penalizing sharp changes. The output displacement fields have subpixel precision and they are used to construct a super-resolution HDR image.

Ferradans et al. [FBPC12] find dense correspondence of input images in the radiance domain with respect to L_{ref}. In order to detect the mismatches in the estimated flow fields $\mathbf{u}_n$, the input images are warped using the estimated fields and the absolute difference map $M_n(p)$ of each pixel $L_n(p)$ is calculated. Instead of applying a fixed threshold to $M_n(p)$, its histogram is modeled as a mixture of Gaussians. $\mathbf{u}_n(p)$ is detected as a mismatch if the following is true:

$$|M_n(p) - \mu| > \beta\sigma, \qquad (27)$$

where μ and σ are the mean and standard deviation of the most probable Gaussian fit, respectively, and β is a user-defined factor. The pixel intensities corresponding to the flow vectors causing the mismatch are assigned zero weight in HDR reconstruction. The information from the remaining pixels in each L_n are fused in the gradient domain.

Jinno and Okuda [JO12] use a novel weighting function which has significantly smaller overlap between the contribution of input LDR images to the radiance domain. The proposed method assumes that the global alignment is already performed. Displacement, occlusion, and saturation regions are modeled as Markov Random Fields $\mathbf{d} = \{d(p)\}$, $\mathbf{o} = \{o(p)\}$ and $\mathbf{s} = \{s(p)\}$ respectively, where $p \in \Lambda$ and $\Lambda = \{(i,j)|(i,j) \in \mathbb{R}^2\}$ is the discrete sampling lattice. $\mathbf{o}$ and $\mathbf{s}$ are binary random fields. The optimal $\mathbf{d}$, $\mathbf{o}$, and $\mathbf{s}$ are found by minimizing the following energy function:

$$(\mathbf{d}^*, \mathbf{o}^*, \mathbf{s}^*) = \underset{\mathbf{d}, \mathbf{o}, \mathbf{s}}{\operatorname{argmin}} \{U(L_i|\mathbf{d}, \mathbf{o}, \mathbf{s}, L_j) + U(\mathbf{d}|\mathbf{o}, \mathbf{s}, L_j)$$
$$+ U(\mathbf{o}|\mathbf{s}, L_j) + U(\mathbf{s}|L_j)\}. \qquad (28)$$

$U(L_i|\mathbf{d}, \mathbf{o}, \mathbf{s}, L_j)$ measures the accuracy of the motion

estimation, ignoring saturation and occlusion regions. $U(\mathbf{d}|\mathbf{o},\mathbf{s},L_j)$ measures the smoothness of displacement vectors. $U(\mathbf{o}|\mathbf{s},L_j)$ and $U(\mathbf{s}|L_j)$ penalize isolated small regions in saturation and occlusion maps. The resulting motion estimates, saturation and occlusion regions are used to combine the input images.

Hafner et al. [HDW14] propose an energy-minimization approach which simultaneously calculates HDR irradiance together with the displacement fields. The displacement fields have sub-pixel accuracy, similar to Zimmer et al. [ZBW11]. The energy function is defined as:

$$\mathcal{E}(H,\mathbf{W}) = \int_{\Omega} \left(\sum_{i=1}^{N} M_i + \alpha \cdot \sum_{i=1}^{N} S_{\mathbf{u}_i} + \beta \cdot S_H \right) dx, \quad (29)$$

where Ω represents the rectangular image domain, $\mathbf{W} = \{\mathbf{u}_1,...,\mathbf{u}_N\}$ is the set of displacement fields, M_i is the data term measuring the difference between the predicted and actual pixel values, $S_{\mathbf{u}_i}$ is the spatial smoothness term of the displacement field and S_H is the spatial smoothness term of the irradiance map. A coarse-to-fine pyramid structure is used in the minimization to avoid local minima.

3.4.2. Patch-based

Patch-based algorithms aim to recover or reconstruct the potential ghost regions in the output image by transferring information from a subset of input images which are determined via a patch-based matching strategy. Empirically, the methods described in this class seem to generate the highest quality outputs. However, due to the intensive searching and patching operations, they are computationally the most costly as well.

Menzel and Guthe [MG07] introduce a motion compensation method which addresses both the camera and the scene motion. Their study takes into account the parallax and occlusion effects caused by the camera movement as well. The proposed method is limited to three input images $\{L_1,L_2,L_3\}$. First, L_1 and L_3 are aligned to L_2 using a method called hierarchical block matching (HBM). The HBM operation is based on the motion estimation of each macroblock M in L_{ref}, maximizing the cross correlation between each L_i and the reference image L_{ref}. Instead of using a fixed macroblock size, the matching procedure is performed in an hierarchical manner. The pixel-wise displacements are estimated using bilinear interpolation of the smallest macroblocks in the hierarchy. Then the HDR image is synthesized using Equation 3. Possible mismatches and ghost regions are detected using cross correlation. The weights, α, of the pixels in these regions are set to zero.

The algorithm of Park et al. [POK*11] operates on two uncompressed Bayer RAW images R_s and R_l with different exposures. In their study, Park et al. handle the unsuccessful alignments and color artifacts caused by over and underexposed pixels in or around the moving objects. From the two

input images R_s with short exposure and R_l with long exposure, R_s is selected as the reference image. In R_s, instead of clipping the under-exposed pixels with low SNR, the wavelet denoising method of Yoo et al. [YLC*10] is applied. After the denoising, the exposure of R_s is normalized using histogram matching. The exposure-normalized R_s and R_l are spatially aligned using hierarchical block matching. Ghost regions are detected in two steps. In the first step, pixel intensity differences are thresholded using a large threshold value, which provides a high confidence detection. In the second step, a region growing operation is applied to the initial ghost regions using a smaller threshold value. Therefore, the ghost regions with lower confidence are ignored if they are not spatially connected to any high confidence detection. The ghost regions in R_l are patched using R_s. During HDR reconstruction, R_l and R_s are registered in two steps. In the first step, background alignment is performed to handle the camera movement. In the second step, foreground alignment is performed which addresses object movements.

Zheng et al. [ZLZR11] introduce a method which consists of a pixel-level movement detection followed by a hybrid patch-based scheme. First, the inconsistent pixels which may cause ghosting artifacts are identified using the method of Li et al. [LRZ*10] (see above) with pairwise comparison of the subsequent LDR images. Among the compared pair of images, if the IMF is reliable, the inconsistent pixels are reconstructed by transferring pixel intensities using the IMF. The IMF is assumed to be unreliable if the pixel intensity values are closer to 0 or 255. In those cases, the absolute derivative of the camera response function tends to get very large and the IMF does not provide a one-to-one mapping between pixel intensities of different exposures. If the IMF is unreliable, then iterative block-based patching is performed instead of the pixel-wise patching. The block-based patching searches for the best matching block in a predefined search window in both pairs of the compared LDR images. The resulting patched LDR images are fused to an HDR image using the method of Debevec and Malik [DM97].

Hu et al. [HGP12] introduce a homography-based patching approach to handle the scene movement. First, they find the dense correspondence between each input image and the reference image based on HaCohen et al. [HSGL11]. Then, the IMF is estimated for each color channel by fitting cubic Hermite splines using RANSAC [FB81] to the observed pixel intensity pairs. The estimated IMF is used to find motion regions where the pixel intensities are not consistent with the IMF. The gaps formed after the forward warping operation and the motion regions detected using the IMF are filled by defining local homographies between the images using RANSAC. Due to the irregular shapes of the gaps, a rectangular bounding box is defined around each gap which typically covers some of the pixels previously synthesized using forward warping as well. To test for the robustness of the defined homography, the normalized cross-correlation is measured between the pixels which are already synthe-

sized and the corresponding pixels from the source images. A high normalized cross-correlation indicates a consistent patch, in which case the patching is performed successfully. Otherwise, only the reference image is used to fill the gaps. Resulting images are fused with the method of Mertens et al. [MKVR07].

Sen et al. [SKY*12] propose a PatchMatch [BSFG09] based energy minimization approach for HDR reconstruction. The proposed approach is designed for LDR images with linearized CRFs. The energy function is in the following form:

$$\mathcal{E}(H) = \sum_{p \in pixels} \Big[\alpha_{ref}(p) \cdot (E_{ref}(p) - H(p))^2 \\ + (1 - \alpha_{ref}(p)) \cdot \mathcal{E}_{MBDS}(H|L_1,...,L_N) \Big],$$

(30)

where $\mathcal{E}_{MBDS}$ is the Multisource Bidirectional Similarity (MBDS) measure which is an extension to BDS introduced by Simakov et al. [SCSI08]. In case of poorly exposed pixels, α function decreases the weight of the information transferred from the reference image, while the weight of the second term is increased, which transfers information from the other input images. Subsequently, the output HDR image is aligned to the reference LDR image and it contains maximum amount of information from the reference image if the pixels are well-exposed. Instead of solving for output HDR image directly, auxiliary images are obtained using the search and vote scheme of Simakov et al. [SCSI08]. Later these auxiliary images are merged to obtain an intermediate HDR image. The auxiliary images are iteratively initialized and updated until convergence using the intermediate HDR image and the search and vote approach. This procedure is performed over multiple scales.

Orozco et al. [OMLV12] presents a method which consists of both ghost detection and image registration steps. In the ghost detection step, the detection algorithms of Pece and Kautz [PK10], Jacobs et al. [JLW08], Sidibe et al. [SPS*09] and Grossberg et al. [GN03] are compared and it was found that the IMF based ghost detection of Grossberg et al. is the most accurate. In the image registration phase, an intensity-based method without feature detection is employed. The image with the best exposure is selected as the reference image. A bounding box is fitted around the previously detected motion regions. Next, the region in each bounding box is registered by translation and rotation to the reference image. The Sum of Squared Distances (SSD), Normalized Cross Correlation (NCC), Mutual Information (MI) and Median Bitmap Difference (MBD) are compared as a similarity measure for the registration. The authors state that NCC has the best computational cost and performance. In order to speed up the process, the registration is performed using the pyramid structure of the images, from coarse to fine resolution. However, since the registration applies only translational and rotational transformations, more complex motions caused by objects with deformable bodies are not handled.

Hu et al.'s more recent work [HGPS13] proposes another PatchMatch [BSGF10] based HDR reconstruction algorithm with energy minimization. Among the input LDR images, the one with the largest number of well-exposed pixels is selected as L_{ref}. Next, for each input LDR image L_i, a latent image T_i is synthesized. Latent images are similar to L_{ref} where it is well-exposed. In under- or over-exposed regions, a matching patch is found using the PatchMatch algorithm in other input images. Using the matching patches and the intensity mapping function obtained with the histogram-based method of Grossberg and Nayar [GN03], the latent images are obtained by minimizing the following energy function:

$$\mathcal{E}(\mathbb{T}, \tau, \mathbf{u}) = C_r(\mathbb{T}, L_{ref}, \tau) + C_t(\mathbb{L}, \mathbb{T}, \mathbf{u}),$$

(31)

where $\mathbb{L}$, $\mathbb{T}$ and $\mathbf{u}$ are the sets of input images, latent images and coordinate mappings to matching patches, respectively. The C_r and C_t terms measure the radiometric and the texture consistencies between the reference image and the input images, respectively. As opposed to Sen et al. [SKY*12], Hu et al. does not require the CRFs of the input images to be linear. In one comparison study [TAEE14], it is observed that Hu et al. was more successful at producing noise-free outputs whereas Sen et al. was better at preserving texture details.

In their more recent work, Zheng et al. [ZLZ*12,ZLZ*13] formulate the patching operation as an optimization problem minimizing the following function for each input image L_n:

$$\sum_{p \in M_n} \|\nabla \Lambda_{n,r(n)}(L_n(p)) - \nabla L_{r(n)}(p)\|_2,$$

(32)

where M_n is the set of pixels affected by motion, ∇ is the gradient operator, $\Lambda_{n,r(n)}$ is the intensity mapping function from L_n to $L_{r(n)}$ and $r(n)$ is the index of the reference image for L_n.

3.5. HDR Video Deghosting

The methods introduced in this section are specially crafted for HDR video deghosting. Although they share some common approaches with the previous HDR deghosting methods, such as the optical-flow and patch-based registration operations, they have some distinct properties which are only applicable to videos.

Kang et al. [KUWS03] proposed an optical-flow based image warping method which is applied to the LDR frames captured using temporal exposure bracketing to produce an HDR video. Similar to [Bog00], they apply a global affine transformation followed by a local optical-flow based correction. For the motion estimation, a variant of the Lucas-Kanade [LK*81] is used, which works on the Laplacian pyramid representation as proposed by Bergen et al. [BAHH92]. With three consecutive exposures L_{n-1}, L_n, L_{n+1} and L_n being the target, L_{n-1} and L_{n+1} are unidirectionally warped to L_n. If L_n is ill-exposed, the unidirectional optical-flow estimation is unreliable. In that case, an interpolated frame I_n is created using only L_{n-1} and L_{n+1}

in bidirectional motion estimation. In order to align I_n with L_n, the authors use a hierarchical homography-based registration.

Sand and Teller [ST04] propose an approach for the spatio-temporal alignment of the frames from two videos of a dynamic scene. Their alignment algorithm provides both an optical-flow field and a temporal offset between matching frames. The flow field is estimated by finding pixel-wise correspondences using Harris corner detector [HS88] and assigning a weight to each one of them, depending on the correspondence quality. The weights consist of two terms; namely, a pixel matching probability and a motion consistency probability. The pixel matching probability is calculated by comparing the pixel intensity of the primary frame with the minimum and maximum intensities observed in the 3×3 neighborhood of the corresponding pixel in the secondary frame. The second term of the weight, the motion consistency probability, is based on the difference between the observed motion vector and the motion vector predicted using adaptive locally weighted regression.

Mangiat and Gibson [MG10] propose a deghosting method designed for HDR video reconstruction from the frames of an LDR video with alternating short and long exposures similar to Kang et al. [KUWS03]. The motion estimation process begins with normalizing short exposure frame L_s using:

$$\tilde{L}_l = g^{-1}(g(L_s) - \ln \Delta t_s + \ln \Delta t_l), \tag{33}$$

where $g = \ln f^{-1}$, L_l is the long exposure frame, Δt_s and Δt_l are the corresponding exposure times. The authors state that an optical-flow approach similar to Kang et al. [KUWS03] is not suitable here due to amplified noise and the possibility of large displacements. Instead, block-based forward and backward motion vectors for the current frame L_n is estimated using the previous frame L_{n-1} and the next frame L_{n+1}. Motion vectors are calculated using Enhanced Predictive Zonal Search (EPZS) in H.264 JM Reference software using Sum of Absolute Differences (SAD) matching measure. After forward and backward motion estimation, similar to Kang et al. [KUWS03], bidirectional motion estimation is performed. Bidirectional motion estimation provides the motion vectors for the saturated blocks in L_n. The obtained motion fields are refined using a method similar to the pixel-level refinement of Matsushita et al. [MOG*06]. In order to prevent the boundary artifacts around blocks, cross-bilateral filtering is applied to the output HDR frames. The outputs may contain some registration artifacts on the saturated moving objects, which is addressed by the authors' more recent work [MG11].

Castro et al. [CCCV11] propose an algorithm which is suitable for portable platforms with limited computational resources. The input of the algorithm is the frames of an LDR video with 0, +1 and −1 EV. After a photometric calibration step, for each frame in the triplet, the alignment of the remaining 2 frames is performed using the method of Ward [War03]. The object motion is addressed by calculating the variance of each pixel in three exposures. If the variance is low, it is assumed that the set of three frames is free of motion and all of them are used during radiance map construction. Otherwise, a larger weight is given to a single frame.

Chapiro et al. [CCV11] provide an application of the exposure fusion [MKVR07] to HDR videos by introducing a fourth term to the weighting function. This term takes lower values in the presence of motion, which is detected by using the total absolute difference of pixel blocks as a measure.

Kalantari et al. [KSB*13] introduce a patch-based HDR synthesis method. Similar to Kang et al. [KUWS03] and Mangiat and Gibson [MG10], the inputs are frames of an LDR video with periodically alternating exposures. There is no reduction in the number of frames in the HDR reconstruction process. For each frame L_n, different exposures are constructed using the information from temporally neighbor frames L_{n-1} and L_{n+1}. The proposed framework is an extension of Sen et al. [SKY*12] to video with temporal coherence. This is done by replacing Bidirectional Similarity (BDS) term with Temporal Bidirectional Similarity (TBDS), which measures BDS of L_n with L_{n-1} and L_{n+1}. In order to accelerate the search and vote procedure, the patch searches in TBDS are constrained around an initial motion estimation which is based on planar model for global motion estimation and optical flow for local motion estimation. Although perceptually insignificant, the authors state that use of motion estimation and optical flow may sometimes result in artifacts around motion boundaries such as blurring and partially disappearing object parts.

4. The Experiment

Our subjective evaluation methodology is similar to the approach followed by Rubinstein et al. [RGSS10] where the authors conducted a *pairwise comparison* experiment via a web-based interface. The set of evaluated deghosting algorithms were: (**A**) Grosch [Gro06], (**B**) Khan et al. [KAR06], (**C**) Sen et al. [SKY*12], (**D**) Silk and Lang [SL12], (**E**) Hu et al. [HGPS13], (**F**) a simple baseline deghosting algorithm discussed in the next subsection, and (**G**) no deghosting as a control condition.

Among these algorithms, (**A**) was selected as a relatively simple older-generation deghosting algorithm. (**B**) was selected as the representative of the category which aims to completely eliminate moving objects. (**C**) and (**E**) were selected as they represent highly sophisticated state-of-the-art algorithms. (**D**) was selected as a relatively simple but more recent algorithm. (**F**), which is described in the next subsection, was selected as a simple baseline algorithm. Finally, (**G**) which represents no deghosting was selected as a control condition to assess the reliability of the subjective experiment.

The implementations of **A**, **C**, **D**, and **E** were available by the authors of the algorithms (for **A** as an executable kindly provided to us by the author). For **D** the fluid motion (FM) outputs were tested since the number of input images was relatively low and they were dynamic throughout whole acquisition process. Although an implementation of **B** is provided as part of open-source software Hugin-2013.0.0 [Hug], we reimplemented it in MATLAB as the Hugin results did not replicate the same results for the images used by Khan et al. in their original paper. The simple deghosting algorithm, **F**, was also implemented by the current authors.

4.1. Simple Deghosting Algorithm

The simple deghosting algorithm is our implementation of a simple ghost detection followed by a Laplacian pyramid [BA83] blending operation for merging images. The input images were generated from the RAW files with linear camera response functions (CRF). For each pair of input images L_i and L_j, a ghost bitmap G_{ij} is obtained with:

$$G_{ij}(p) = \frac{|L_i(p)\frac{\Delta t_j}{\Delta t_i} - L_j(p)|}{L_j(p)} > 0.1, \qquad (34)$$

which marks a pixel in the bitmap if there is more than 10% deviation in the predicted pixel intensity value. Only well-exposed pixels are used in this ghost detection operation. The final ghost detection mask is obtained by merging all bitmaps with logical-OR operation. In the final step, HDR image H is obtained by using the method of Debevec and Malik [DM97] and the regions which are marked in the ghost detection mask are filled using the pixel values from only the middle exposed image in the radiance domain. Laplacian pyramid [BA83] is used to avoid seams between the regions taken only from the middle exposure and the neighboring pixels computed from multiple images. Note that this algorithm does not involve global exposure registration. However, in case of camera movement the MTB algorithm [War03] could be used to align the exposures in a pre-processing step.

4.2. Benchmark Dataset

For the experiments, we acquired images of 10 different scenes with different characteristics. The acquisition settings and the image properties of the scenes used in the experiment are given in Table 2. The stacks are taken with a tripod and no preliminary global registration is applied to the images. Each scene consisted of 3 LDR images with ± 1 EV difference. Each input image was resized to 1024×683 dimensions for computational considerations. For any specific scene, only the exposure time was varied among the exposures while the ISO setting and F-number parameters were fixed. The input image with 0 EV was used as the reference exposure if a reference image was required by the

tested algorithms. All images were captured in the RAW format. Subsequently, RAW images were converted to 8-bit LDR images with linear camera response function (CRF) using the DCRaw [Cof] software. The output of HDR images were tonemapped using the photographic tonemapping operator [RSSF02] for visualization purposes. A representative image for each of these scenes is shown in Figure 3.

We give priority to cover the most frequently observed real-world ghosting scenarios as much as possible in the set of test scenes. Both indoor and outdoor scenes, small and large object displacements, deformable and non-deformable motion patterns as well as different types of moving objects, lighting conditions and noise levels are mostly represented in the dataset. However, the number of the test scenes is open for growth and we are planning to extend the dataset further in our future studies, focusing on the test cases where it is critical for the algorithms to recover the full dynamic range of saturated motion regions.

4.3. Experimental Setup

As mentioned above, the experimental design was pairwise comparisons similar to Rubinstein et al. [RGSS10]. However, as deghosting algorithms operate on an exposure stack, rather than a single image as in image retargeting, the participants were presented with 3 exposures on the left side of the screen. To maximize image size, only one exposure was shown in high resolution. By hovering the mouse over the thumbnails at the top the participants could view each input exposure in high resolution. Furthermore, by hovering the mouse over the exposures a zoomed-in view of the region under the mouse pointer was presented as an overlay in a small window. A pair of deghosting results were presented on the right side of the screen which could be switched and zoomed-in similar to the input exposures. To indicate their preference, the participants first selected the thumbnail that corresponds to the preferred result and then clicked the submit button at the bottom of the page (see Figure 4).

For each participant, the experiment started with a short warm-up session comprised of 3 comparisons during which the responses were not recorded. During the actual experiment, each participant compared 60 pairs of images. The exact phrase used in the comparison page was "Please select the image that you think is the better deghosting result created from the multiple exposures." The progress bar at the bottom of the page showed the participants' progress. All participants performed the experiment using their own computers (as in a crowd-sourcing study) and were able to finish the experiment within 30 minutes.

The first page of the web-interface briefly informed the participants about the HDR deghosting problem and what was expected of them. It also collected information about the participants' age, gender, and familiarity with computer graphics and image processing. In total, 63 participants (13F

| (a) Cafe | (b) Candles | (c) FastCars | (d) Flag | (e) Flag |

| (f) Gallery2 | (g) LibrarySide | (h) Shop1 | (i) Shop2 | (j) WalkingPeople |
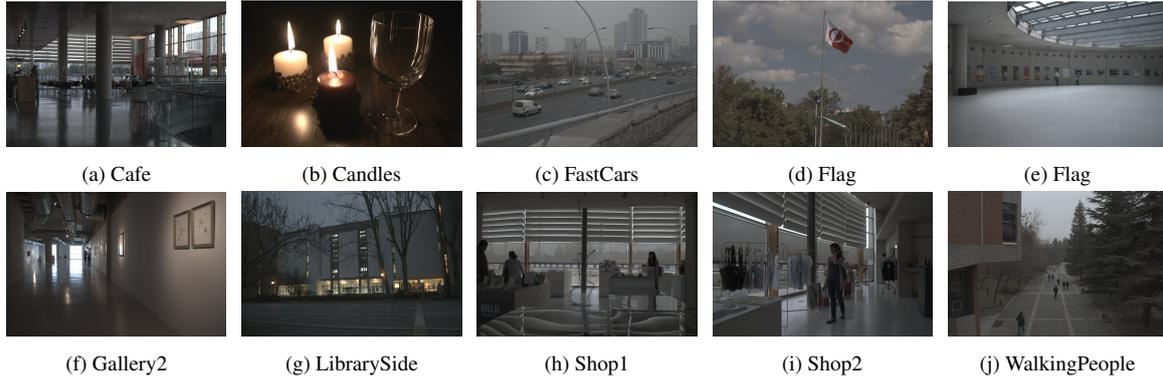
Figure 3: A representative image for each scene

Table 2: Acquisition settings and image properties for the scenes used in the experiment.

Scene	Camera	Exposure Time	ISO	F-number	Properties
Cafe	Nikon D5100	1/250	320	5.0	Indoor scene with multiple moving human subjects
Candles	Canon 550D	1/128	6400	4.6	Indoor scene with low-lighting conditions including moving light sources and high noise
FastCars	Canon 550D	1/256	1600	5.0	Outdoor scene including non-deformable body motion with large spatial displacements
Flag	Canon 550D	1/256	100	11.3	Outdoor scene with deformable-body motion
Gallery1	Nikon D5100	1/250	250	4.0	Indoor scene with moving human subjects on dim background
Gallery2	Nikon D5100	1/250	2500	3.5	Indoor scene with moving human subjects on strongly-lit background
LibrarySide	Canon 550D	1/100	6400	3.5	Outdoor scene with low-lighting conditions including moving people in strongly-lit environment
Shop1	Nikon D5100	1/250	320	5.6	Indoor scene with reflections and moving human subjects on strongly-lit background
Shop2	Nikon D5100	1/250	320	7.1	Indoor scene with a single moving human subjects
WalkingPeople	Canon 550D	1/256	200	4.6	Outdoor scene with moving human subjects with occlusion

and 50M) finished the experiment from the beginning to the end, and their data was used in the subsequent analysis. Of the 63 participants, 33 indicated they work in the field of computer graphics and image processing, 14 indicated they consider this field as a hobby, and 16 indicated that they do not have any specific interest in the field. The participants ages were distributed between 21 and 50 with the mode age being 25 and the mean 31.

We note that the total number of comparison pairs was equal to $10 \times \binom{7}{2} = 210$. Each participant evaluated a random selection of 60 pairs among these. To ensure that each pair was evaluated equal number of times, we saved information about how many times a pair was compared. When a new participant started the experiment, the least frequently compared 60 pairs were selected in random order. This ensured that when 7 participants completed the experiment, each of the 210 pairs was compared exactly twice ($7 \times 60 = 210 \times 2$). Thus, after all 63 participants finished the experiment, each possible pair was compared exactly 18 times.

4.4. Data Analysis

In a paired comparison test, each participant makes a binary choice in each one of the possible pairs of items. With t items, there are $\binom{t}{2}$ pairs to compare. The results of the comparisons are represented by an aggregate preference matrix $A = [a_{ij}]$ where a_{ij} is the number of times item i is
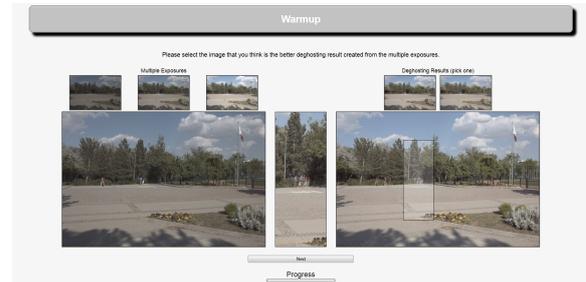
Figure 4: Screenshot of the pairwise comparison task.

preferred over algorithm j. The probability of item i being preferred over item j is π_{ij}. The mean of the probability of an algorithm i being preferred over other algorithms is:

$$\pi_i = \frac{1}{t-1} \sum_{j=1, j \neq i}^{t} \pi_{ij}. \quad (35)$$

An estimate of π_i is given by:

$$\pi_i = \frac{a_i}{n(t-1)}, \quad (36)$$

where n is the number of comparisons per item and $a_i = \sum_{j=1}^{t} a_{ij}$. In order to analyze the significance of the

Table 3: The aggregate preference matrices for each scene used in the experiment. Each cell shows the number of times the row algorithm is preferred over the column algorithm. Please refer to text for the algorithm labels. The statistical similarity groups are indicated in the last column with lower group numbers corresponding to higher preference.

Cafe	A	B	C	D	E	F	G	Sum	Group(s)
A	0	18	3	17	2	2	17	59	2
B	0	0	4	18	0	2	13	37	3
C	15	14	0	18	13	16	16	92	1
D	1	0	0	0	0	0	0	1	4
E	16	18	5	18	0	8	16	81	1
F	16	16	2	18	10	0	18	80	1
G	1	5	2	18	2	0	0	28	3

Candles	A	B	C	D	E	F	G	Sum	Group(s)
A	0	18	1	17	1	0	10	47	3
B	0	0	0	0	1	1	2	4	4
C	17	18	0	18	15	17	18	103	1
D	1	18	0	0	0	1	0	20	4
E	17	17	3	18	0	15	18	88	1, 2
F	18	17	1	17	3	0	17	73	2
G	8	16	0	18	0	1	0	43	3

FastCars	A	B	C	D	E	F	G	Sum	Group(s)
A	0	18	1	17	2	1	17	56	3
B	0	0	1	8	2	1	12	24	4
C	17	17	0	18	4	10	17	83	1, 2
D	1	10	0	0	0	2	8	21	4
E	16	16	14	18	0	15	17	96	1
F	17	17	8	16	3	0	17	78	2
G	1	6	1	10	1	1	0	20	4

Flag	A	B	C	D	E	F	G	Sum	Group(s)
A	0	18	12	17	10	6	15	78	2
B	0	0	0	1	0	1	7	9	4
C	6	18	0	18	9	3	15	69	2
D	1	17	0	0	2	0	15	35	3
E	8	18	9	16	0	2	16	69	2
F	12	17	15	18	16	0	18	96	1
G	3	11	3	3	2	0	0	22	3, 4

Gallery1	A	B	C	D	E	F	G	Sum	Group(s)
A	0	15	6	14	6	9	16	66	2
B	3	0	1	9	2	2	16	33	3
C	12	17	0	16	6	10	16	77	1, 2
D	4	9	2	0	0	1	12	28	3, 4
E	12	16	12	18	0	13	16	87	1
F	9	16	8	17	5	0	17	72	1, 2
G	2	2	2	6	2	1	0	15	4

Gallery2	A	B	C	D	E	F	G	Sum	Group(s)
A	0	16	2	17	10	6	18	69	2
B	2	0	1	17	4	2	17	43	3
C	16	17	0	18	17	11	16	95	1
D	1	1	0	0	2	1	3	8	4
E	8	14	1	16	0	0	16	55	2,3
F	12	16	7	17	18	0	17	87	1
G	0	1	2	15	2	1	0	21	4

LibrarySide	A	B	C	D	E	F	G	Sum	Group(s)
A	0	16	12	18	11	16	18	91	1
B	2	0	2	17	1	1	13	36	3
C	6	16	0	17	13	17	16	85	1
D	0	1	1	0	1	1	3	7	4
E	7	17	5	17	0	14	16	76	1
F	2	17	1	17	4	0	14	55	2
G	0	5	2	15	2	4	0	28	3

Shop1	A	B	C	D	E	F	G	Sum	Group(s)
A	0	15	9	17	6	7	18	72	1
B	3	0	4	12	4	4	15	42	2
C	9	14	0	18	5	10	18	74	1
D	1	6	0	0	1	2	13	23	3
E	12	14	13	17	0	11	17	84	1
F	11	14	8	16	7	0	17	73	1
G	0	3	0	5	1	1	0	10	3

Shop2	A	B	C	D	E	F	G	Sum	Group(s)
A	0	12	2	13	2	1	16	46	2
B	6	0	4	14	3	2	17	46	2
C	16	14	0	14	15	7	17	83	1
D	5	4	4	0	1	3	17	34	2
E	16	15	3	17	0	11	17	79	1
F	17	16	11	15	7	0	17	83	1
G	2	1	1	1	1	1	0	7	3

WalkingPeople	A	B	C	D	E	F	G	Sum	Group(s)
A	0	16	8	17	2	12	17	72	2,3
B	2	0	2	16	1	1	16	38	4
C	10	16	0	17	5	16	18	82	1,2
D	1	2	1	0	0	1	6	11	5
E	16	17	13	18	0	15	18	97	1
F	6	17	2	17	3	0	17	62	3
G	1	2	0	12	0	1	0	16	5

scores, we used the statistical data analysis method of Starks and David [SD61] which tests the following null hypothesis:

$$H_0 : \pi_i = \pi_j, \forall i, j. \qquad (37)$$

A special case of the test given by Durbin [Dur51] expects that if H_0 is true (i.e. all compared items are alike), the following D value follows approximately χ^2 distribution with $t - 1$ degrees of freedom:

$$D = \sum_{i=1}^{t} d_i^2 = \frac{4}{nt} \sum_{i=1}^{t} (a_i - \bar{a})^2,$$

$$= \frac{4}{nt} \sum_{i=1}^{t} a_i^2 - \frac{1}{4} t n^2 (t-1)^2. \qquad (38)$$

Using the formula of Durbin, if D is greater than the critical value χ_α^2 for a selected significance level α, it is possible to reject H_0. For a significance value of $\alpha = 0.05$, the corresponding $\chi_{0.05}^2 = 12.592$. In our experiment, it was possible to reject H_0 for each scene with D values greater than 100

Table 4: Total aggregate preference matrix of the participants in the subjective experiment and algorithm scores consisting of total number of preferences

	A	B	C	D	E	F	G	Sum
A	0	162	56	164	52	60	162	656
B	18	0	19	112	18	17	128	312
C	124	161	0	172	102	117	167	843
D	16	68	8	0	7	12	77	188
E	128	162	78	173	0	104	167	812
F	120	163	63	168	76	0	169	759
G	18	52	13	103	13	11	0	210

Table 5: Ranks and significant groups of the algorithms according to the scores (a_i)

Group	Algorithms
1	C - Sen et al. [SKY*12] (843), E - Hu et al. [HGPS13] (812)
2	F - SimpleDG (759)
3	A - Grosch [Gro06] (656)
4	B - Khan et al. [KAR06] (312)
5	G - NoDG (210), D - Silk and Lang [SL12] (188)

(with the minimum *D* as 118.5 for the *Shop2* scene). Rejection of H_0 allows one to perform pairwise comparison tests in order to group the algorithms into statistical significance groups. For two scores a_i and a_j, Starks and David [SD61] calculate the smallest amount of statistically significant difference required as:

$$m_c = \lceil 1.96(0.5nt)^{0.5} + 0.5 \rceil. \tag{39}$$

If $|a_i - a_j| \geq m_c$ is satisfied, it is possible to conclude that there is a statistically significant difference between the scores of compared items with a significance level of $\alpha = 0.05$. In this study, we found that $m_c = 17$ for each scene and $m_c = 50$ for aggregate results. The significant groups determined using these m_c values are given in Table 3 for each scene and in Table 5 for total aggregate preference matrix.

5. Results

In this section, we first present the results of the subjective experiment and interpret them based on the outputs generated by the algorithms. Next, we discuss the runtime performance of the algorithms. Finally, we investigate whether the observed findings can be correlated with two objective metrics.

5.1. Experimental Results

The results of the subjective experiment for each scene is given in Table 3. In this table, each matrix represents the results for a single scene. The cell values indicate how many times the row algorithm was preferred over the column algorithm. The last column for each scene indicates the statistical similarity groups with lower numbered groups corresponding to higher preference. The aggregate results obtained by accumulating the preference matrices for each scene is given in Table 4. The statistical similarity groups for the aggregate results are separately presented in Table 5 for clarity purposes.

From these results we can make the following observations. Sen et al.'s [SKY*12] and Hu et al.'s [HGPS13] methods are clear winners based on the aggregate rankings. These

methods also outperform all other methods for each scene with a few exceptions. In general, Sen et al.'s method was selected as the best algorithm for all scenes except the Flag scene. Hu et al.'s method was also in the first group except the Gallery2 and Flag scenes. When these methods were not the winner, they ranked the second. This suggests that these two patch-based algorithms are quite stable with respect to changing scene contents and different types of ghosting artifacts. In general, both methods appeared to be artifact-free. However, it was observed that they may produce outputs that have slightly less contrast compared to Grosch [Gro06] and the simple deghosting algorithm discussed in Section 4.1. This may explain their second ranking for the Flag scene as all four methods produced artifact-free images but the latter two produced higher contrast (please refer to supplementary materials as these differences are best observed in high resolution images).

An interesting observation is that the simple deghosting algorithm (**F**) explained in Section 4.1 performed relatively well in the experiment; it was placed by itself in the second group. In fact, the per-scene results suggest that many times this simple algorithm was ranked in the first significance group (for Cafe, Gallery2, Shop1, Flag, Shop2, and Gallery1). The worst result for this algorithm was observed in WalkingPeople where the simple algorithm was ranked as the third. This suggests that, when it comes to deghosting, a simple solution may sometimes outperform more sophisticated algorithms assuming that the exposures are captured using a tripod or registered during a preprocessing step (this simple method generally outperformed Grosch, Khan et al., and Silk and Lang's algorithms). This may be attributed to the fact that the simple method does not generate any additional artifacts which are sometimes observed in more sophisticated algorithms.

Grosch's method [Gro06] seems to have the highest amount of variance between the rankings in different scenes. While it performs very well (ranked in the first group) for LibrarySide, Shop1, and Flag scenes it performs the worst for the Candles scene. In other scenes, it occupies the second and third rankings. This variation suggests that this algorithm's results are highly sensitive to scene content and the types of ghosting artifacts that are present. Figure 5 il-
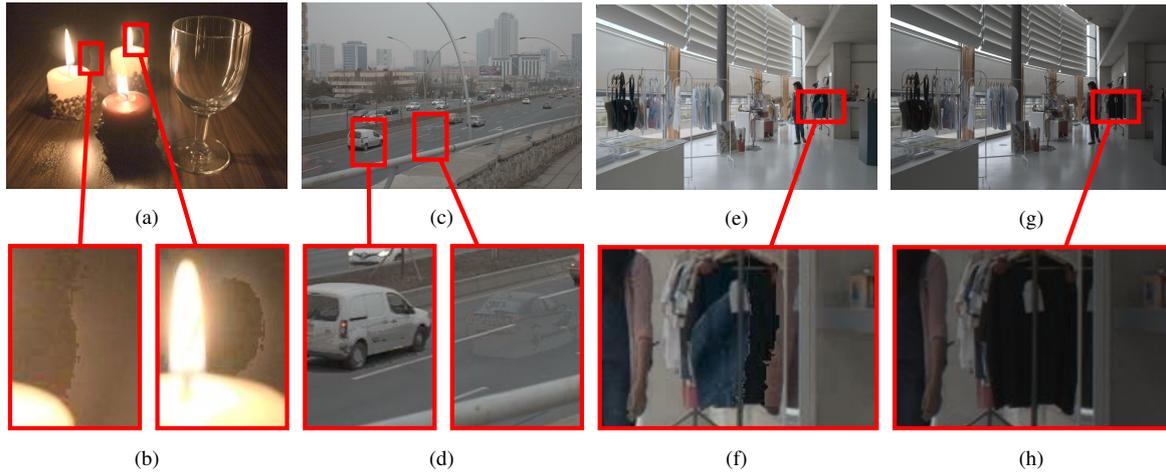
Figure 5: Outputs of Grosch [Gro06] for selected scenes. (a), (c), and (e) show the results for Candles, FastCars and Shop2 scenes, (g) shows one of the input LDR exposures for Shop2, (b), (d), (f) and (h) magnify the problematic regions. In (b), the sharp transition between the source images is easily observable which is attributed to the lack of a smooth blending operation. In (d), some parts of the car are replaced with the background on the left and some parts of another car remain on the right due to the color similarity between the object and the background. In (f), the presence of the black colored cloth (underexposed pixels) and the movement at the same regions result in incorrect filling of the region, leaving visible boundaries. The same region of a single input exposure is provided in (h) for reference.

lustrates some of the artifacts created by this algorithm and discusses their causes.

Khan et al.'s method [KAR06] performed relatively poorly in the experiment with also a high degree of variation. It performed the worst for the Candles scene which depicts a low-light environment with high noise. It is quite possible that the amount of noise present in the exposures interfered with the algorithm's weight computation. For a few scenes such as Shop1 and Shop2, this algorithm performed relatively well, occupying the second ranking. For the remaining scenes, Khan et al.'s algorithm occupied the third and fourth rankings. As such, this algorithm also exhibited a high degree of variation across scenes. One possible reason for the low performance of this algorithm may be attributed to the fact that the exposure stacks were comprised of only three images. Because this algorithms assigns weights to each pixel by considering its similarity to the 3D neighborhood around it, using only three images may have given rise to a too small neighborhood. Some of the artifacts with their possible causes are presented in Figure 6.

The worst performing deghosting algorithm was found to be Silk and Lang's method [SL12]. Overall, this algorithm was preferred the fewest number of times in pairwise comparisons, receiving a score even lower than the no deghosting condition. We attribute this to the artifacts produced by this algorithm. In the outputs of Silk and Lang, it is observed that, especially in low-lit surfaces, the outputs have black regions (even if these regions are completely static). In addition, possibly due to the blending operation used, the transition in the

super-pixel boundaries may become very sharp or they may produce color discontinuities as shown in Figure 7.

In overall, the obtained rankings give confidence about the reliability of the subjective experiment. Sen et al. and Hu et al., being very similar algorithms, shared the first ranking. The no deghosting control condition and Silk and Lang's [SL12] algorithm occupied the last position, a finding that is expected from the artifacts in their outputs. Simple deghosting and Grosch's methods [Gro06] received rankings that are similar to each other. This is also expected as both algorithms are similar but the simple deghosting includes a Laplacian blending stage whereas Grosch's algorithm simply uses pixels from the reference exposure.

5.2. Runtime Performance

Running times of each algorithm is provided in Table 6. The running times were obtained by measuring the CPU time used by each algorithm on a computer platform with Intel Core i7-3770 CPU @ 3.40 GHz, 8 GB RAM and NVIDIA GeForce GT 630 GPU.

For Sen et al. [SKY*12], the "normal" quality setting was used. For Khan et al. [KAR06], the iteration count was set to 10. However, we report the running times per-iteration as we observed that $4 - 5$ iterations were sufficient for convergence, most of the time. All algorithm implementations were in MATLAB, excluding Grosch [Gro06] which was kindly provided to us by the author in executable format. The fast
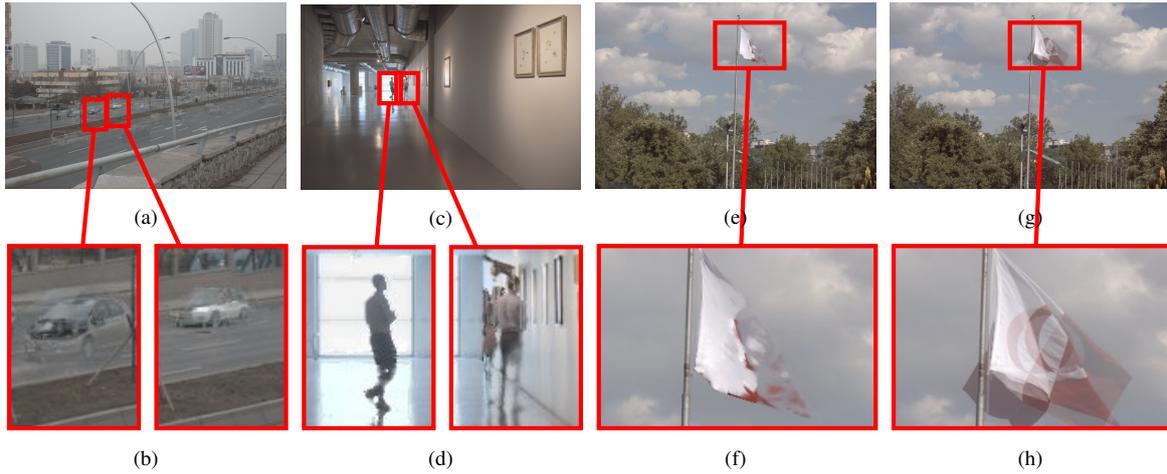
Figure 6: Outputs of Khan et al. [KAR06] in (a) FastCars, (c) Gallery2 and (e) Flag scenes. The problematic regions are magnified in (b), (d), (f) and (h). In FastCars scene, the critical assumption of the algorithm does not hold. A vehicle in the scene takes the position of another vehicle from the previous frame; therefore, the majority of the exposures do not capture the background in these regions, which is required for a correct pixel-weighting operation. This situation results in even increased amount of weights where the pixels are affected by the motion in (b). Since there is not a semantic constraint in the pixel down-weighting operations, two copies of the same person appears in the Gallery2 (d) scene. (f) shows the output of Khan et al. in the presence of deformable body motion. The overlapping parts of the moving object are kept whereas other parts are cleared by the algorithm giving rise to a broken appearance. In (h), the same region is shown with no deghosting operation for reference.
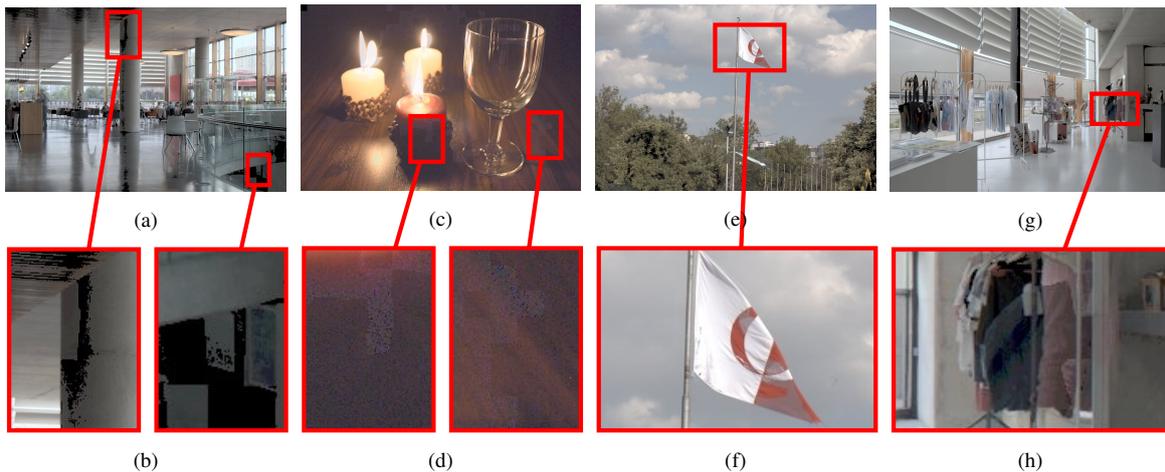
Figure 7: Outputs of Silk and Lang [SL12] for (a) Cafe, (c) Candles, (e) Flag, and (g) Shop2 scenes. In the outputs of the algorithm some of the regions become black, especially if the they have low pixel-intensity values in the input images. In (b), two such regions are shown. In (d), two regions are shown where the boundaries of super-pixels are visually noticeable. (e) and (f) shows the observed color artifacts after the blending operation. (h) shows a region in Shop2 scene where multiple sources are used to fill an underexposed region.

running times can be attributed to its being native code and the algorithm's utilization of the GPU.

From Table 6 we can observe that Sen et al.'s and Hu et al.'s methods take about $3 - 4$ minutes to process an exposure stack comprised of 3 exposures with each exposure 1024 × 683 resolution. Khan et al.'s running times take about 10 minutes for a single iteration of the algorithm. The other methods are much faster, especially Grosch's method, producing results in about a second (including disk IO times).

Table 6: Algorithm running times in seconds. Please refer to text for details.

	Grosch	Khan et al.	Sen et al.	Silk and Lang	Hu et al.	SimpleDG	NoDG
Cafe	1.06	615.03	218.08	24.95	208.28	8.80	3.89
Candles	1.09	624.05	308.77	8.17	299.34	7.03	2.30
FastCars	1.05	620.86	184.77	11.77	239.42	7.72	2.70
Flag	1.06	616.94	161.09	12.64	232.52	6.36	3.41
Gallery1	1.02	613.92	173.36	18.44	221.47	6.42	2.86
Gallery2	0.98	611.28	218.83	14.66	243.59	6.45	2.05
LibrarySide	1.09	616.89	238.58	18.09	215.77	7.58	2.72
Shop1	1.00	614.42	203.77	11.02	215.91	6.78	2.53
Shop2	0.95	613.14	199.63	10.61	218.44	6.88	3.45
WalkingPeople	1.09	617.96	190.97	12.92	208.83	6.88	3.47
Average	1.04	616.45	209.78	14.33	230.36	7.09	2.94

5.3. Objective Evaluation

Conducting a large-scale subjective experiment for comparing HDR deghosting algorithms is a challenging task. With new algorithms being proposed on a regular basis, the findings obtained from a subjective experiment can be quickly outdated. Therefore, it is important to develop objective metrics which can be used to quickly and quantitatively evaluate the performance of newly proposed algorithms. Here, we share the results for two simple metrics that seemed plausible to be among the first attempts in this direction. To measure how well these metrics can evaluate deghosting performance, we calculated Spearman's rank correlation (ρ) between the scores from the subjective experiment and the scores of these metrics which produces a score in range $[-1, 1]$.

Deblurring metric. The first metric that we tested was Liu et al.'s [LWC*13] no-reference metric for evaluating the quality of motion deblurring. The motivation for using this metric is that blurring artifacts and ghosting artifacts are somewhat similar and therefore a metric for the former may be used to predict the quality for the latter. The results of this metric are provided in Table 7. In this table, larger values indicate higher predicted quality. Spearman's rank correlation coefficients between the rankings according to this metric and the subjective experiment are given in the last column of this table. In general, it can be seen that there is low correlation, a finding which indicates that Liu et al.'s metric is not suitable for evaluating HDR deghosting algorithms.

Dynamic range. We experimented with the dynamic range as a simple objective metric, knowing that even if this metric gives high correlations it is not likely to be a sufficient metric on its own as one can easily create an extremely high dynamic range image full of artifacts. However, the motivation for this metric was that, in general, as the ghosting artifacts appear due to blending between different objects they tend to result in reduced dynamic range (e.g., consider a sharp candle against a candle that has ghosting artifacts). Therefore,

a higher dynamic range may indicate a better deghosting result. The results of this metric and its correlation with the results of the subjective experiment is shown in Table 8. The last column of this table reports Spearman's rank correlation coefficient. We can observe that there is a higher degree of correlation between the metric's results and the experimental results. However, for some scenes there is zero or negative correlation. As such, it can be argued that dynamic range has a promise but it must be either used as an auxiliary metric beside a more advanced metric or customized, for example, to only include pixels in ghost regions.

6. Conclusions and Future Work

In this study, an extensive review of the state of the art in HDR deghosting literature and a novel hierarchical taxonomy of deghosting algorithms is provided. Using this taxonomy, we classified approximately 50 HDR deghosting algorithms. We also discussed the distinguishing characteristics of the algorithms within each class. This survey also included a subjective experiment involving 63 participants to compare 6 deghosting algorithms with varying degrees of complexity. These algorithms were evaluated using 10 scenes with different characteristics. This benchmark dataset, and the proposed evaluation framework, is suitable to be used in future evaluation studies as well. Finally, we experimented with the possibility of using objective metrics to quantitatively compare HDR deghosting algorithms. To this end, two metrics, one designed for evaluating deblurring algorithms, and one based on a simple dynamic range measure were utilized. Given the increased activity in this field, we identify objective evaluation of ghost removal techniques as the most important future work direction.

Table 7: The objective scores obtained using Liu et al. [LWC*13] (larger is better) and Spearman's rank correlation ρ with the subjective experiment scores. The average correlation ρ̄ = −0.001.

	Grosch	**Khan et al.**	**Sen et al.**	**Silk and Lang**	**Hu et al.**	**SimpleDG**	**NoDG**	**ρ**
Cafe	5.68	4.60	4.86	8.44	4.68	7.31	6.33	-0.46
Candles	-0.31	-0.74	-0.42	-1.03	-1.14	-0.12	-1.17	0.25
FastCars	1.36	1.54	1.20	0.72	1.36	1.91	1.00	0.45
Flag	0.03	-0.73	-0.32	-0.01	-0.38	0.46	-0.50	0.88
Gallery1	1.29	0.48	1.02	1.06	1.39	1.87	1.42	0.04
Gallery2	0.77	-0.28	-0.37	0.37	-0.19	0.53	0.06	-0.11
LibrarySide	4.90	4.13	3.82	5.42	3.69	4.38	3.55	-0.04
Shop1	9.61	8.29	8.45	9.69	8.77	10.31	9.24	-0.18
Shop2	6.88	6.16	6.38	7.42	6.06	7.07	6.33	-0.02
WalkingPeople	1.61	1.94	1.72	3.04	1.53	2.03	1.82	-0.82

Table 8: Dynamic Range computed in log_{10} units with 1% of the lightest and darkest pixels excluded and Spearman's rank correlation ρ with the subjective experiment scores. The average correlation ρ̄ = 0.45.

	Grosch	**Khan et al.**	**Sen et al.**	**Silk and Lang**	**Hu et al.**	**SimpleDG**	**NoDG**	**ρ**
Cafe	2.63	2.55	2.61	2.60	2.46	2.50	2.47	0.00
Candles	2.66	2.43	3.03	2.67	2.85	2.66	2.65	0.77
FastCars	1.12	1.10	1.18	1.10	1.15	1.17	1.10	0.85
Flag	1.40	1.49	1.50	1.49	1.46	1.44	1.45	-0.52
Gallery1	1.59	1.49	1.59	1.56	1.56	1.57	1.55	0.52
Gallery2	2.41	2.24	2.56	2.14	2.31	2.31	2.29	0.90
LibrarySide	1.78	1.79	1.93	1.60	1.87	1.75	1.76	0.64
Shop1	2.20	2.04	2.39	2.00	2.33	2.43	2.10	0.75
Shop2	2.68	2.45	2.72	2.89	2.64	2.70	2.55	0.29
WalkingPeople	1.94	2.06	2.07	1.83	2.02	2.05	2.05	0.25

References

[ADA*04] AGARWALA A., DONTCHEVA M., AGRAWALA M., DRUCKER S., COLBURN A., CURLESS B., SALESIN D., COHEN M.: Interactive digital photomontage. *ACM Trans. on Graphics 23*, 3 (2004), 294–302. 5

[Aky11] AKYÜZ A. O.: Photographically guided alignment for HDR images. *Eurographics-Areas Papers* (2011), 73–74. 4, 5

[ALKC11] AN J., LEE S. H., KUK J. G., CHO N. I.: A multi-exposure image fusion algorithm without ghost effect. In *Acoustics, Speech and Signal Processing (ICASSP), IEEE Int. Conf. on* (2011), IEEE, pp. 1565–1568. 4, 9

[ASS*10] ACHANTA R., SHAJI A., SMITH K., LUCCHI A., FUA P., SÜSSTRUNK S.: *SLIC superpixels*. Tech. rep., EPFL Technical Report, 2010. 6

[BA83] BURT P. J., ADELSON E. H.: The laplacian pyramid as a compact image code. *Communications, IEEE Transactions on 31*, 4 (1983), 532–540. 11, 15

[BADC11] BANTERLE F., ARTUSI A., DEBATTISTA K., CHALMERS A.: *Advanced High Dynamic Range Imaging: Theory and Practice*, first edition ed. CRC Press (AK Peters), Natick, MA, 2011. 1

[BAHH92] BERGEN J. R., ANANDAN P., HANNA K. J., HIN-

GORANI R.: Hierarchical model-based motion estimation. In *Computer Vision ECCV'92* (1992), Springer, pp. 237–252. 13

[BK93] BURT P. J., KOLCZYNSKI R. J.: Enhanced image capture through fusion. In *Computer Vision, Fourth Int. Conf. on* (1993), IEEE, pp. 173–182. 2

[BL03] BROWN M., LOWE D. G.: Recognising panoramas. In *Proceedings of the 9th IEEE Int. Conf. on Computer Vision - Volume 2* (Washington, DC, USA, 2003), ICCV, IEEE Computer Society. 5

[Bog00] BOGONI L.: Extending dynamic range of monochrome and color images through fusion. In *Pattern Recognition, 15th Int. Conf. on* (2000), vol. 3, IEEE, pp. 7–12. 4, 11, 13

[BSFG09] BARNES C., SHECHTMAN E., FINKELSTEIN A., GOLDMAN D.: PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Trans. on Graphics 28*, 3 (2009), 24. 13

[BSGF10] BARNES C., SHECHTMAN E., GOLDMAN D. B., FINKELSTEIN A.: The generalized PatchMatch correspondence algorithm. In *Computer Vision–ECCV*. Springer, 2010, pp. 29–43. 13

[BSL*11] BAKER S., SCHARSTEIN D., LEWIS J., ROTH S., BLACK M. J., SZELISKI R.: A database and evaluation methodology for optical flow. *Int. Journal of Computer Vision 92*, 1 (2011), 1–31. 3

[BVZ01] BOYKOV Y., VEKSLER O., ZABIH R.: Fast approximate energy minimization via graph cuts. *Pattern Analysis*

and Machine Intelligence, IEEE Transactions on 23, 11 (2001), 1222–1239. 6, 9

[Can03] CANDOCIA F. M.: Simultaneous homographic and comparametric alignment of multiple exposure-adjusted pictures of the same scene. *Image Processing, IEEE Transactions on 12*, 12 (2003), 1485–1494. 4

[CCCV11] CASTRO T., CHAPIRO A., CICCONET M., VELHO L.: Towards mobile hdr video. *Proceedings of the Eurographics-Areas Papers* (2011), 75–76. 4, 14

[CCV11] CHAPIRO A., CICCONET M., VELHO L.: Filter based deghosting for exposure fusion video. In *ACM SIGGRAPH Posters* (2011), ACM, p. 33. 4, 14

[CH06] CERMAN L., HLAVAC V.: Exposure time estimation for high dynamic range imaging with hand held camera. In *Proc. of Computer Vision Winter Workshop, Czech Republic* (2006), Citeseer. 4, 5

[Cof] COFFIN D.: DCRaw: Decoding raw digital photos in linux. Accessed: February 2014. URL: http://www.cybercom.net/~dcoffin/dcraw/. 15

[CS07] COMBY F., STRAUSS O.: Using quasi-continuous histograms for fuzzy main motion estimation in video sequence. *Fuzzy Sets and Systems 158*, 5 (2007), 475–495. 6

[DM97] DEBEVEC P. E., MALIK J.: Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th Annual Conf. on Computer Graphics and Interactive Techniques* (1997), SIGGRAPH, ACM Press/Addison-Wesley Publishing Co., pp. 369–378. 2, 5, 9, 12, 15

[Dur51] DURBIN J.: Incomplete blocks in ranking experiments. *British Journal of Statistical Psychology 4*, 2 (1951), 85–90. 17

[EUS06] EDEN A., UYTTENDAELE M., SZELISKI R.: Seamless image stitching of scenes with large motions and exposure differences. In *Computer Vision and Pattern Recognition, IEEE Computer Soc. Conf. on* (2006), vol. 2, pp. 2498–2505. 4, 5

[FB81] FISCHLER M. A., BOLLES R. C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM 24*, 6 (1981), 381–395. 5, 12

[FBPC12] FERRADANS S., BERTALMÍO M., PROVENZI E., CASELLES V.: Generation of HDR images in non-static conditions based on gradient fusion. In *VISAPP* (2012), pp. 31–37. 4, 11

[FLW02] FATTAL R., LISCHINSKI D., WERMAN M.: Gradient domain high dynamic range compression. *ACM Trans. on Graphics 21*, 3 (2002), 249–256. 8

[GAW*10] GRANADOS M., AJDIN B., WAND M., THEOBALT C., SEIDEL H.-P., LENSCH H.: Optimal HDR reconstruction with linear digital cameras. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conf. on* (2010), IEEE, pp. 215–222. 2

[GG07] GEVREKCI M., GUNTURK B. K.: On geometric and photometric registration of images. In *Acoustics, Speech and Signal Processing, IEEE Int. Conf. on* (2007), vol. 1, IEEE. 4, 5

[GGC*09] GALLO O., GELFAND N., CHEN W.-C., TICO M., PULLI K.: Artifact-free high dynamic range imaging. In *Computational Photography (ICCP), IEEE Int. Conf. on* (2009), IEEE, pp. 1–7. 4, 8

[GKTT13] GRANADOS M., KIM K. I., TOMPKIN J., THEOBALT C.: Automatic noise modeling for ghost-free HDR reconstruction. *ACM Trans. on Graphics 32*, 6 (2013), 201. 4, 10

[GN03] GROSSBERG M. D., NAYAR S. K.: Determining the camera response from images: What is knowable? *Pattern Analysis and Machine Intelligence, IEEE Transactions on 25*, 11 (2003), 1455–1467. 2, 7, 13

[Gro06] GROSCH T.: Fast and robust high dynamic range image generation with camera and object movement. *Vision, Modeling and Visualization, RWTH Aachen* (2006), 277–284. 4, 7, 9, 14, 18, 19, 21

[GSL08] GRANADOS M., SEIDEL H.-P., LENSCH H.: Background estimation from non-time sequence images. In *Proceedings of Graphics Interface* (2008), Canadian Information Processing Society, pp. 33–40. 4, 6

[HDW14] HAFNER D., DEMETZ O., WEICKERT J.: Simultaneous HDR and optic flow computation. In *Pattern Recognition (ICPR 2014), Proc. 22nd Int. Conf. on* (2014), IEEE. 4, 12

[HG11] HOSSAIN I., GUNTURK B. K.: High dynamic range imaging of non-static scenes. In *IS&T/SPIE Electronic Imaging* (2011), Int. Society for Optics and Photonics. 4, 11

[HGP12] HU J., GALLO O., PULLI K.: Exposure stacks of live scenes with hand-held cameras. In *Computer Vision–ECCV*. Springer, 2012, pp. 499–512. 4, 12

[HGPS13] HU J., GALLO O., PULLI K., SUN X.: HDR deghosting: How to deal with saturation. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conf. on. IEEE* (2013). 3, 4, 13, 14, 18

[HLL*10] HEO Y. S., LEE K. M., LEE S. U., MOON Y., CHA J.: Ghost-free high dynamic range imaging. In *Computer Vision–ACCV*. Springer, 2010, pp. 486–500. 4, 9

[HS88] HARRIS C., STEPHENS M.: A combined corner and edge detector. In *Alvey vision Conf.* (1988), vol. 15, Manchester, UK, p. 50. 14

[HSGL11] HACOHEN Y., SHECHTMAN E., GOLDMAN D. B., LISCHINSKI D.: Non-rigid dense correspondence with applications for image enhancement. *ACM Trans. on Graphics 30*, 4 (2011), 70. 12

[HTM14] HADZIABDIC K. K., TELALOVIC J. H., MANTIUK R.: Expert evaluation of deghosting algorithms for multi-exposure high dynamic range imaging. In *Second Int. Conf. and SME Workshop on HDR imaging (HDRi2014)* (2014). 3

[Hug] HUGIN: Panorama photo stitcher. Accessed: September 2014. URL: http://hugin.sourceforge.net/. 15

[IJLP11] IM J., JANG S., LEE S., PAIK J.: Geometrical transformation-based ghost artifacts removing for high dynamic range image. In *Image Processing (ICIP), 18th IEEE Int. Conf. on* (2011), IEEE, pp. 357–360. 5

[ILP11] IM J., LEE S., PAIK J.: Improved elastic registration for removing ghost artifacts in high dynamic imaging. *Consumer Electronics, IEEE Transactions on 57*, 2 (2011), 932–935. 4, 5

[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. 4, 7, 8, 13

[JO12] JINNO T., OKUDA M.: Multiple exposure fusion for high dynamic range image acquisition. *Image Processing, IEEE Transactions on 21*, 1 (2012), 358–365. 4, 11

[KAR06] KHAN E. A., AKYÜZ A. O., REINHARD E.: Ghost removal in high dynamic range images. In *Image Processing, IEEE Int. Conf. on* (2006), IEEE, pp. 2005–2008. 4, 6, 14, 18, 19, 20

[KHC*06] KAO W.-C., HSU C.-C., CHEN L.-Y., KAO C.-C., CHEN S.-H.: Integrating image fusion and motion stabilization

for capturing still images in high dynamic range scenes. *Consumer Electronics, IEEE Transactions on 52*, 3 (2006), 735–741. 4, 7

[KHTM13] KARADUZOVIC-HADZIABDIC K., TELALOVIC J. H., MANTIUK R.: Comparison of deghosting algorithms for multi-exposure high dynamic range imaging. In *Spring Conf. on Computer Graphics* (2013), ACM, pp. 21–28. 3

[Kov99] KOVESI P.: Image features from phase congruency. *Videre: Jrnl. of computer vision research 1*, 3 (1999), 1–26. 5

[KSB∗13] KALANTARI N. K., SHECHTMAN E., BARNES C., DARABI S., GOLDMAN D. B., SEN P.: Patch-based high dynamic range video. *ACM Trans. Graph. 32*, 6 (2013). 4, 14

[KUWS03] KANG S. B., UYTTENDAELE M., WINDER S., SZELISKI R.: High dynamic range video. *ACM Trans. on Graphics 22*, 3 (2003), 319–325. 4, 13, 14

[LC09] LIN H.-Y., CHANG W.-Z.: High dynamic range imaging for stereoscopic scene representation. In *Image Processing (ICIP), 16th IEEE Int. Conf. on* (2009), pp. 4305–4308. 4, 7

[LK∗81] LUCAS B. D., KANADE T., ET AL.: An iterative image registration technique with an application to stereo vision. In *IJCAI* (1981), vol. 81, pp. 674–679. 13

[LLM14] LEE C., LI Y., MONGA V.: Ghost-free high dynamic range imaging via rank minimization. *Signal Processing Letters, IEEE 21*, 9 (2014), 1045–1049. 4, 10

[Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *Int. journal of computer vision 60*, 2 (2004), 91–110. 5

[LPC11] LEE D.-K., PARK R.-H., CHANG S.: Improved histogram based ghost removal in exposure fusion for high dynamic range images. In *Consumer Electronics (ISCE), IEEE 15th Int. Symposium on* (2011), IEEE, pp. 586–591. 4, 8

[LRZ∗10] LI Z., RAHARDJA S., ZHU Z., XIE S., WU S.: Movement detection for the synthesis of high dynamic range images. In *Image Processing (ICIP), 17th IEEE Int. Conf. on* (2010), IEEE, pp. 3133–3136. 4, 8, 12

[LWC∗13] LIU Y., WANG J., CHO S., FINKELSTEIN A., RUSINKIEWICZ S.: A no-reference metric for evaluating the quality of motion deblurring. *ACM Trans. Graph. 32*, 6 (2013), 175. 3, 21, 22

[MG07] MENZEL N., GUTHE M.: Freehand HDR photography with motion compensation. In *VMV* (2007), pp. 127–134. 4, 12

[MG10] MANGIAT S., GIBSON J.: High dynamic range video with ghost removal. In *SPIE Optical Engineering+ Applications* (2010), Int. Society for Optics and Photonics. 4, 14

[MG11] MANGIAT S., GIBSON J.: Spatially adaptive filtering for registration artifact removal in hdr video. In *Image Processing, 18th IEEE Int. Conf. on* (2011), IEEE, pp. 1317–1320. 14

[MKVR07] MERTENS T., KAUTZ J., VAN REETH F.: Exposure fusion. In *Computer Graphics and Applications, 15th Pacific Conf. on* (2007), IEEE, pp. 382–390. 8, 9, 10, 13, 14

[MMF02] MANN S., MANDERS C., FUNG J.: Painting with looks: Photographic images from video using quantimetric processing. In *Proceedings of the tenth ACM Int. Conf. on Multimedia* (2002), ACM, pp. 117–126. 4

[MN99] MITSUNAGA T., NAYAR S. K.: Radiometric self calibration. In *Computer Vision and Pattern Recognition, IEEE Computer Society Conf. on.* (1999), vol. 1, IEEE. 2

[MOG∗06] MATSUSHITA Y., OFEK E., GE W., TANG X., SHUM H.-Y.: Full-frame video stabilization with motion inpainting. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 28*, 7 (2006), 1150–1163. 14

[MOW∗07] MALM H., OSKARSSON M., WARRANT E., CLARBERG P., HASSELGREN J., LEJDFORS C.: Adaptive enhancement and noise reduction in very low light-level video. In *Computer Vision, IEEE 11th Int. Conf. on* (2007), IEEE, pp. 1–8. 9

[MP94] MANN S., PICARD R. W.: *On Being 'undigital' With Digital Cameras: Extending Dynamic Range By Combining Differently Exposed Pictures*. Tech. Rep. 323, M.I.T. Media Lab Perceptual Computing Section, Boston, Massachusetts, 1994. also appears IS&T's 48th annual conference, Cambridge, Massachusetts, May 1995. 2

[MPC09] MIN T.-H., PARK R.-H., CHANG S.: Histogram based ghost removal in high dynamic range images. In *Multimedia and Expo, IEEE Int. Conf. on* (2009), IEEE, pp. 530–533. 4, 8, 9

[MPC11] MIN T.-H., PARK R.-H., CHANG S.: Noise reduction in high dynamic range images. *Signal, Image and Video Processing 5*, 3 (2011), 315–328. 4, 8, 9

[MTCL12] MOON Y.-S., TAI Y.-M., CHA J. H., LEE S.-H.: A simple ghost-free exposure fusion for embedded HDR imaging. In *Consumer Electronics (ICCE), IEEE Int. Conf. on* (2012), IEEE, pp. 9–10. 4, 9

[MW06] MALM H., WARRANT E.: Motion dependent spatiotemporal smoothing for noise reduction in very dim light image sequences. In *Pattern Recognition, 18th Int. Conf. on* (2006), vol. 3, IEEE, pp. 954–959. 9

[OAG13] OĞUZ AKYÜZ A., GENCTAV A.: A reality check for radiometric camera response recovery algorithms. *Computers & Graphics 37*, 7 (2013), 935–943. 2

[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. 4, 10

[OMLV12] OROZCO R. R., MARTÍN I., LOSCOS C., VASQUEZ P.-P.: Full high-dynamic range images for dynamic scenes. In *SPIE Photonics Europe* (2012), Int. Society for Optics and Photonics. 4, 13

[Par62] PARZEN E.: On estimation of a probability density function and mode. *The annals of mathematical statistics* (1962), 1065–1076. 9

[PF03] PERIASWAMY S., FARID H.: Elastic registration in the presence of intensity variations. *Medical Imaging, IEEE Transactions on 22*, 7 (2003), 865–874. 5

[PH08] PEDONE M., HEIKKILÄ J.: Constrain propagation for ghost removal in high dynamic range images. In *VISAPP* (2008), pp. 36–41. 4, 6

[PK10] PECE F., KAUTZ J.: Bitmap movement detection: HDR for dynamic scenes. In *Visual Media Production (CVMP), Conf. on* (2010), IEEE, pp. 1–8. 4, 8, 13

[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 *Image Processing (ICIP), 18th IEEE Int. Conf. on* (2011), IEEE, pp. 353–356. 4, 12

[R∗56] ROSENBLATT M., ET AL.: Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics 27*, 3 (1956), 832–837. 9

[RBS99] ROBERTSON M. A., BORMAN S., STEVENSON R. L.: Dynamic range improvement through multiple exposures. In *Image Processing, Int. Conf. on* (1999), vol. 3, IEEE, pp. 159–163. 2

[RC01] RICHARD M. M. O. B. B., CHANG M. Y.-S.: Fast digital image inpainting. In *Appeared in the Proceedings of the Int. Conf. on Visualization, Imaging and Image Processing (VIIP)* (2001), pp. 106–107. 9

[RC10] RAMAN S., CHAUDHURI S.: Bottom-up segmentation for ghost-free reconstruction of a dynamic scene from multi-exposure images. In *Proceedings of the Seventh Indian Conf. on Computer Vision, Graphics and Image Processing* (2010), ACM, pp. 56–63. 4, 8, 9

[RC11] RAMAN S., CHAUDHURI S.: Reconstruction of high contrast images for dynamic scenes. *The Visual Computer 27*, 12 (2011), 1099–1114. 4, 9

[RGSS10] RUBINSTEIN M., GUTIERREZ D., SORKINE O., SHAMIR A.: A comparative study of image retargeting. *ACM Trans. Graph. 29*, 6 (2010). 3, 14, 15

[RKC09] RAMAN S., KUMAR V., CHAUDHURI S.: Blind deghosting for automatic multi-exposure compositing. In *SIGGRAPH ASIA Posters* (2009), ACM, p. 44. 4, 8, 9

[RMVS07] RAD A. A., MEYLAN L., VANDEWALLE P., SÜSSTRUNK S.: Multidimensional image enhancement from a set of unregistered and differently exposed images. In *Electronic Imaging* (2007), Int. Society for Optics and Photonics. 4, 5

[RSSF02] REINHARD E., STARK M., SHIRLEY P., FERWERDA J.: Photographic tone reproduction for digital images. *ACM Trans. on Graphics 21*, 3 (2002), 267–276. 15

[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. 1, 4, 7, 8

[SB05] SHIRAHATTI N. V., BARNARD K.: Evaluating image retrieval. In *Computer Vision and Pattern Recognition, IEEE Computer Society Conf. on* (2005), vol. 1, IEEE, pp. 955–961. 3

[SCSI08] SIMAKOV D., CASPI Y., SHECHTMAN E., IRANI M.: Summarizing visual data using bidirectional similarity. In *Computer Vision and Pattern Recognition, IEEE Conf. on* (2008), IEEE, pp. 1–8. 13

[SD61] STARKS T., DAVID H.: Significance tests for paired-comparison experiments. *Biometrika* (1961), 95–108. 17, 18

[SH92] SHAPIRO R. H. L., HARALICK R.: Computer and robot vision. *Reading: Addison-Wesley* (1992). 8

[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 Trans. Graph. 31*, 6 (2012), 203. 3, 4, 13, 14, 18, 19

[SL12] SILK S., LANG J.: Fast high dynamic range image deghosting for arbitrary scene motion. In *Proceedings of Graphics Interface* (2012), Canadian Information Processing Society, pp. 85–92. 4, 6, 8, 14, 18, 19, 20

[SPLC13] SUNG H.-S., PARK R.-H., LEE D.-K., CHANG S.: Feature based ghost removal in high dynamic range imaging. *Int. Journal of Computer Graphics & Animation 3*, 4 (2013). 4, 10

[SPS*09] SIDIBE D. D., PUECH W., STRAUSS O., ET AL.: Ghost detection and removal in high dynamic range images. In *European Signal Processing Conf.* (2009). 4, 6, 13

[SS12] SRIKANTHA A., SIDIBÉ D.: Ghost detection and removal for high dynamic range images: Recent advances. *Signal Processing: Image Communication 27*, 6 (2012), 650–662. 3

[SSM12] SRIKANTHA A., SIDIBÉ D., MÉRIAUDEAU F.: An SVD-based approach for ghost detection and removal in high dynamic range images. In *Pattern Recognition (ICPR), 21st Int. Conf. on* (2012), IEEE, pp. 380–383. 4, 10

[ST04] SAND P., TELLER S.: Video matching. *ACM Trans. on Graphics 23*, 3 (2004), 592–599. 4, 14

[TAEE14] TURSUN O. T., AKYUZ A. O., ERDEM A., ERDEM E.: Evaluating deghosting algorithms for HDR images. In *Signal Processing and Communications Applications Conf. (SIU), 22nd* (2014), IEEE, pp. 1275–1278. 3, 13

[TDSM07] TOMBARI F., DI STEFANO L., MATTOCCIA S.: A robust measure for visual correspondence. In *Image Analysis and Processing, 14th Int. Conf. on* (2007), IEEE, pp. 376–381. 10

[TM98] TOMASI C., MANDUCHI R.: Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth Int. Conf. on* (1998), IEEE, pp. 839–846. 9

[TM07] TOMASZEWSKA A., MANTIUK R.: Image registration for multi-exposure high dynamic range image acquisition. In *15th Int. Conf. Central Europe on Computer Graphics, Visualization and Computer Vision* (2007). 4, 5, 10

[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. 4, 7, 8, 14, 15

[WT13] WANG C., TU C.: An exposure fusion approach without ghost for dynamic scenes. In *Image and Signal Processing, 6th Int. Congress on* (2013), vol. 2, IEEE, pp. 904–909. 4, 10

[WXRL10] WU S., XIE S., RAHARDJA S., LI Z.: A robust and fast anti-ghosting algorithm for high dynamic range imaging. In *Image Processing (ICIP), 17th IEEE Int. Conf. on* (2010), IEEE, pp. 397–400. 4, 9

[Yao11] YAO S.: Robust image registration for multiple exposure high dynamic range image synthesis. In *IS&T/SPIE Electronic Imaging* (2011), Int. Society for Optics and Photonics. 4, 5

[YBFU96] YANG G.-Z., BURGER P., FIRMIN D. N., UNDERWOOD S.: Structure adaptive anisotropic image filtering. *Image and Vision Computing 14*, 2 (1996), 135–145. 9

[YLC*10] YOO Y., LEE K., CHOE W., PARK S., LEE S.-D., KIM C.-Y.: A digital ISO expansion technique for digital cameras. In *IS&T/SPIE Electronic Imaging* (2010), Int. Society for Optics and Photonics. 12

[ZBW11] ZIMMER H., BRUHN A., WEICKERT J.: Freehand HDR imaging of moving scenes with simultaneous resolution enhancement. *Computer Graphics Forum 30*, 2 (2011), 405–414. 3, 4, 11, 12

[ZC10] ZHANG W., CHAM W.-K.: Gradient-directed composition of multi-exposure images. In *Computer Vision and Pattern Recognition, IEEE Conf. on* (2010), IEEE, pp. 530–536. 4, 6, 10

[ZC12a] ZHANG W., CHAM W.-K.: Gradient-directed multiexposure composition. *Image Processing, IEEE Transactions on 21*, 4 (2012), 2318–2323. 4, 6, 10

[ZC12b] ZHANG W., CHAM W.-K.: Reference-guided exposure fusion in dynamic scenes. *Journal of Visual Communication and Image Representation 23*, 3 (2012), 467–475. 4, 10

[ZLZ*12] ZHENG J., LI Z., ZHU Z., WU S., RAHARDJA S.: Patching of moving objects for ghosting-free hdr synthesis. In *ACM SIGGRAPH Posters* (2012), ACM, p. 62. 4, 13

[ZLZ*13] ZHENG J., LI Z., ZHU Z., WU S., RAHARDJA S.: Hybrid patching for a sequence of differently exposed images with moving objects. *Image Processing, IEEE Transactions on 22*, 12 (2013), 5190–5201. 4, 13

[ZLZR11] ZHENG J., LI Z., ZHU Z., RAHARDJA S.: A hybrid patching scheme for high dynamic range imaging. In *Asia Pacific Signal and Information Processing Association Annual Summit and Conf.* (2011). 4, 12

[ZPB07] ZACH C., POCK T., BISCHOF H.: A duality based approach for realtime TV-L 1 optical flow. In *Pattern Recognition*. Springer, 2007, pp. 214–223. 11