THE STATE OF THE ART IN HDR DEGHOSTING AND AN OBJECTIVE HDR
IMAGE DEGHOSTING QUALITY METRIC

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

OKAN TARHAN TURSUN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
COMPUTER ENGINEERING

JANUARY 2016

Approval of the thesis:

THE STATE OF THE ART IN HDR DEGHOSTING AND AN OBJECTIVE HDR IMAGE DEGHOSTING QUALITY METRIC

submitted by **OKAN TARHAN TURSUN** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences**   _____

Prof. Dr. Adnan Yazıcı
Head of Department, **Computer Engineering**   _____

Assoc. Prof. Dr. Ahmet Oğuz Akyüz
Supervisor, **Computer Engineering Department, METU**   _____

Examining Committee Members:

Prof. Dr. Sibel Tarı
Computer Engineering Department, METU   _____

Assoc. Prof. Dr. Ahmet Oğuz Akyüz
Computer Engineering Department, METU   _____

Assist. Prof. Dr. Erkut Erdem
Computer Engineering Department, Hacettepe University   _____

Assist. Prof. Dr. Sinan Kalkan
Computer Engineering Department, METU   _____

Assist. Prof. Dr. Jasminka Hasić-Telalović
Computer Science and Engineering Department, IUS   _____

Date: _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name:　　OKAN TARHAN TURSUN

Signature　　　　　：

ABSTRACT

THE STATE OF THE ART IN HDR DEGHOSTING AND AN OBJECTIVE HDR IMAGE DEGHOSTING QUALITY METRIC

Tursun, Okan Tarhan

Ph.D., Department of Computer Engineering

Supervisor   : Assoc. Prof. Dr. Ahmet Oğuz Akyüz

January 2016, 120 pages

Despite the emergence of new HDR acquisition methods, the multiple exposure technique (MET) is still the most popular one. The application of MET on dynamic scenes is a challenging task due to the diversity of motion patterns and uncontrollable factors such as sensor noise, scene occlusion and performance concerns on some platforms with limited computational capability. Currently, there are already more than 50 deghosting algorithms proposed for artifact-free HDR imaging of dynamic scenes and it is expected that this number will grow in the future. Due to the large number of algorithms, it is a difficult and time-consuming task to conduct subjective experiments for benchmarking recently proposed algorithms. In this thesis, first, a taxonomy of HDR deghosting methods and the key characteristics of each group of algorithms are introduced. Next, the potential artifacts which are observed frequently in the outputs of HDR deghosting algorithms are defined and an objective HDR image deghosting quality metric is presented. It is found that the proposed metric is well correlated with the human preferences and it may be used as a reference for benchmarking current and future HDR image deghosting algorithms.

Keywords: HDR Imaging, Deghosting, Scene Motion Analysis

ÖZ

YDA GÖRÜNTÜ GÖLGELEME GİDERMEDE GELİŞMİŞLİK SEVİYESİ VE YDA GÖRÜNTÜLER İÇİN NESNEL BİR GÖLGELEME GİDERME KALİTE METRİĞİ

Tursun, Okan Tarhan

Doktora, Bilgisayar Mühendisliği Bölümü

Tez Yöneticisi   : Doç. Dr. Ahmet Oğuz Akyüz

Ocak 2016 , 120 sayfa

Son zamanlarda yeni YDA görüntüleme teknikleri ortaya çıkmasına rağmen, çoklu pozlama tekniği (ÇPT) hala en popüler teknik olmaya devam etmektedir. Ancak hareket örüntülerinin çeşitliliği, algılayıcı kaynaklı gürültü, sahnedeki kapamalar ve sınırlı hesaplama kabiliyeti olan platformlardaki performans kaygıları gibi sebeplerle ÇPT tekniğinin hareketli sahnelere uygulanması hala zorlu bir işlemdir. Hareketli sahnelerin hatasız bir şekilde YDA görüntülenmesi için literatürde önerilmiş olan halihazırda 50'den fazla gölgeleme giderme algoritması bulunmakta ve bu sayının gelecekte artacağı öngörülmektedir. Önerilmiş olan algoritmaların fazlalığı nedeniyle, yeni algoritmaları karşılaştırmak üzere öznel deney yapmak zor ve zaman alıcı bir yöntem haline gelmiştir. Bu tezde, ilk olarak, YDA gölgeleme giderme algoritmaları için bir sınıflandırma sunulmakta ve her sınıfın anahtar özelliklerinden bahsedilmektedir. Ardından, YDA gölgeleme algoritmalarının çıktılarında en sık gözlemlenen görsel hatalar tanımlanmakta ve nesnel bir YDA gölgeleme giderme kalite metriği sunulmaktadır. Önerilen metriğin insan tercihleriyle anlamlı korelasyona sahip olduğu ve algoritmaların değerlendirilmesinde bir referans olarak kullanılabileceği gözlemlenmiştir.

Anahtar Kelimeler: YDA Görüntüleme, Gölgeleme Giderme, Sahne Hareket Analizi

To my family

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

BDS   Bidirectional Similarity

BMD   Bitmap Movement Detection

CIFT   Contrast-Invariant Feature Transform

CPU   Central Processing Unit

CRF   Camera Response Function

DG   Deghosting

DRIM   Dynamic Range Independent Metric

DSLR   Digital Single-Lens Reflex

EPZS   Enhanced Predictive Zonal Search

ER   Elastic Registration

EV   Exposure Value

FM   Fluid Motion

FR   Full Reference

FSIM   Feature Similarity

GPU   Graphics Processing Unit

HBM   Hierarchical Block Matching

HDR-VDP   High Dynamic Range Visible Difference Predictor

HDRI   High Dynamic Range Imaging

HDR   High Dynamic Range

IMF   Intensity Mapping Function

IO   Input Output

ITU-R   ITU Radiocommunication Sector

ITU   International Telecommunication Union

LDR   Low Dynamic Range

LOOCV   Leave-One-Out Cross-Validation

MBDS   Multisource Bidirectional Similarity

MBD   Median Bitmap Difference

MET   Multiple Exposures Technique

MRF	Markov Random Field
MTB	Median Threshold Bitmap
NaN	Not a Number
NCC	Normalized Cross Correlation
NR	No Reference
PC	Phase Congruency
PDF	Probability Density Function
PSNR	Peak Signal-to-Noise Ratio
PWD	Pairwise Downweighting
QCH	Quasi Continuous Histograms
RAM	Random Access Memory
RANSAC	Random Sample Consensus
RGB	Red Green Blue
RR	Reduced Reference
SAD	Sum of Absolute Differences
SIFT	Scale-Invariant Feature Transform
SLIC	Simple Linear Iterative Clustering
SNR	Signal-to-Noise Ratio
SSD	Sum of Squared Differences
SSIM	Structural Similarity
SURF	Speeded Up Robust Features
SVD	Singular Value Decomposition
TBDS	Temporal Bidirectional Similarity
UDQM	Unified Deghosting Quality Metric
UI	Uncertainty Image
VDM	Visual Discrimination Model
VDP	Visible Differences Predictor
VIF	Visual Information Fidelity
VI	Variance Image
ZNCC	Zero-mean Normalized Cross Correlation

CHAPTER 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 [108].

1.1 Problem Definition

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, methods that are collectively called as multiple exposure techniques (MET) (see Figure 1.1). 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 [7].

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 ex-

1

(a) L_1 (EV -4)　　(b) L_2 (EV -3)　　(c) L_3 (EV -2)　　(d) L_4 (EV -1)　　(e) L_5 (EV 0)

(f) L_6 (EV +1)　　(g) L_7 (EV +2)　　(h) L_8 (EV +3)　　(i) L_9 (EV +4)　　(j) I
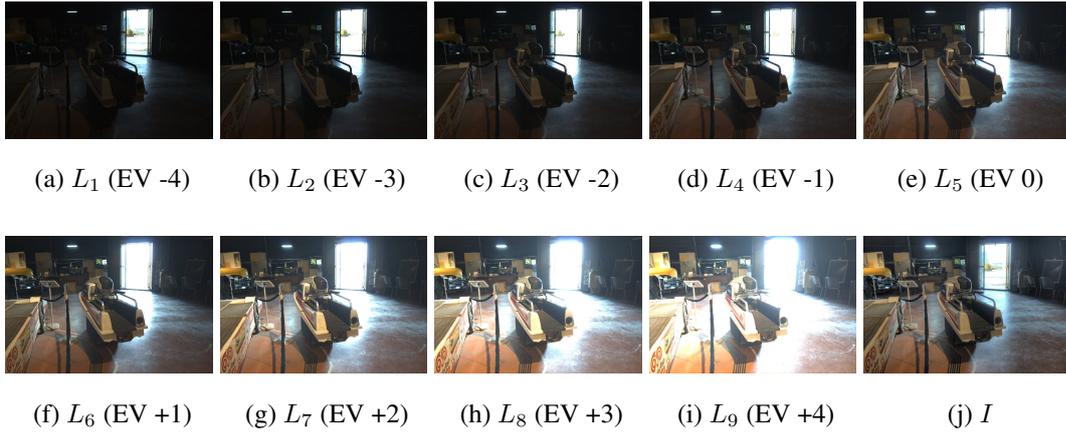
Figure 1.1: Input images $L_1, ..., L_9$ (a-i) and HDR image I (j) (tonemapped) obtained using MET.

(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 1.2: Different types of ghost artifacts.

postures. This is due to the fact that the ensuing HDR image reconstruction process simply 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

Table 1.1: Notation used in this dissertation

$L_1, L_2, ..., L_N$	Input LDR images
L_{ref}	Reference LDR image
$L_n(p)$	Pixel intensity at position (p) in image L_n
I	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
$w_n(p)$	Weight of pixel $L_n(p)$
$E_1, E_2, ..., E_N$	Input LDR image in radiance domain (Eq. 1.2)

Figure 1.2).

It is possible to formalize this notion as follows (see Table 1.1 for the terminology used in this chapter). 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 1.1):

$$L(p) = f(E(p) \cdot \Delta t), \tag{1.1}$$

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:

$$E(p) = \frac{f^{-1}(L(p))}{\Delta t}. \tag{1.2}$$

Most of the time, f is not known but can be recovered using various techniques [16, 25, 39, 82, 90, 96, 110]. 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 $I(p)$ can be computed as:

$$I(p) = \frac{\sum_{n=1}^{N} w_n(p) \frac{f^{-1}(L_n(p))}{\Delta t_n}}{\sum_{n=1}^{N} w_n(p)}, \tag{1.3}$$

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 [25], recent research has shown that other parameters such as the camera noise must be taken into account to determine an optimal weighting function [34].

The critical assumption of Equation 1.3 is that all input images $L_1, ..., L_N$ measure the same scene radiance value for each pixel position p:

$$\frac{f^{-1}(L_n(p))}{\Delta t_n} = \frac{f^{-1}(L_m(p))}{\Delta t_m} \quad \forall n, m, p. \tag{1.4}$$

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

The main requirement of a pixel measuring the same irradiance in all input exposures necessitates that the camera and the captured scene remain static throughout the capture process. While stabilizing a camera can be achieved by using a tripod, ensuring a static scene is much more difficult as most real-world scenes contain dynamic objects. Many deghosting algorithms have been proposed to address this problem ranging from simple alignment methods to sophisticated computer vision algorithms. To this date, more than 50 deghosting algorithms have been proposed.

As in all fields, the proliferation of these algorithms gave rise to subjective experiments that aim to evaluate their performance [42, 43, 44, 136]. However, subjective comparisons of HDR deghosting algorithms is problematic for several reasons. First, ideally the comparison medium must be an HDR display [114], as otherwise some artifacts may be lost or new artifacts may be generated during tone mapping. Secondly, the comparison task is challenging as participants need to compare a stack of LDR images with one or more deghosted images. Finally, the findings of subjective experiments become outdated as new algorithms are being proposed on a rapid basis.

In order to overcome these problems, there is a clear need to define objective metrics to compare HDR deghosting algorithms. A number of quality assessment metrics have been proposed for HDR images [108, Chapter 10] [5, 64, 83, 106, 124]. However, none of these metrics are suitable for evaluating deghosting artifacts.

1.2 Contributions and Outline of the Thesis

The first goal of this dissertation is to conduct an extensive and up-to-date survey and classification of the state of the art in the field of HDR deghosting. To this end, this work includes approximately 50 HDR deghosting methods grouped into a taxonomy based on the approaches they follow. This survey and the taxonomy are presented in Chapter 2.

Due to the increasing number of deghosting algorithms that are being proposed each year, there is a growing need to evaluate these methods systematically and 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 [6], image retrieval [119], and image retargeting [112]. Thus, the second goal of this dissertation is to present new benchmark datasets and the results of subjective experiments aimed to compare several deghosting algorithms. The results of this comparison is provided in Chapter 3.

Third, this work addresses the requirement for objective evaluation of HDR image deghosting algorithms. In Chapter 4, the types of most common deghosting artifacts are identified and objective HDR image deghosting quality metrics are introduced for detecting these artifacts. It is shown that the objective metrics are well correlated with the results of two subjective experiments. Moreover, the individual objective metrics are merged and a unified deghosting quality metric obtained, which has even better correlation with the participant preferences. An application of the objective metric to hybrid deghosting is demonstrated for getting a high quality deghosting result from two input deghosting algorithms.

Fourth, the strengths and weaknesses of different classes of deghosting algorithms and challenges in designing and evaluating deghosting algorithms are discussed in Chapter 5. Finally, a summary of the works done and the future research directions are provided in Chapter 6.

CHAPTER 2

THE STATE OF THE ART IN HDR DEGHOSTING

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.

Srikantha and Sidibé [126] 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. [43] proposed a methodology for evaluating deghosting algorithms and compared the methods of Sen et al. [115] and Zimmer et al. [155] 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 [44], the authors extended their comparison to include the algorithm of Hu et al. [51] and performed an expert evaluation.

In order to organize and highlight the similarities and differences between existing deghosting algorithms, here a taxonomy of approximately 50 methods is presented 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 2.1 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.

2.1 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.

8

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	
[81]	[62]	[61]	[32] [89]	[11]	[86]	[60]
[17]	[100]	[38]	[88] [92]	[49]	[97]	[113]
[140]	[36]	[57]	[105] [127]	[155]	[152]	[79]
[19]	[120]	[71]	[103] [151]	[29]	[95]	[18]
[27]	[121]-PWD	[108]	[70] [94]	[58]	[50]	[20]
[33]	[149]	[99]	[142] [129]	[45]	[115]	[59]
[133]	[150]	[68]	[47] [35]		[51]	
[102]		[121]-FM	[4] [137]		[153]	
[54]			[104] [67]		[154]	
[3]						
[144]						

Figure 2.1: Taxonomy of HDR motion compensation methods

In their early work, Mann et al. [81] 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 [17] introduces another method based on comparametric equations similar to Mann et al.'s method [81], 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 al-

gorithm. Since this optimization is nonlinear with multiple local minima, the initial parameter set is initialized by pair-wise 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 [140] 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:

$$M_i(p) = \begin{cases} 1, \text{if } L_i(p) < \mu_i \\ 0, \text{otherwise} \end{cases} \tag{2.1}$$

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 [19] 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. [27] 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 [14]. Next, the input images are mapped to the radiance domain. In this procedure, a pre-calibrated 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. [2]. 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 [33] 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 [63]. Next, the input images are spatially registered by feature matching using RANSAC.

Tomaszewska and Mantiuk [133] aim to correct misalignments due to the camera motion by estimating a general planar homography using SIFT [74] features and RANSAC [30]. 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. [102] start with estimating the CRF, f, based on Debevec et al. [25]. 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 [19], 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. [53] propose an algorithm based on the elastic registration (ER) method of Periaswamy et al. [101]. 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 [54] 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 [3] 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 [144] provides another method based on the use of phase congruency (PC) images. Different from Gevrekci and Gunturk [33], 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.

2.2 Moving Object Removal

The main objective of the algorithms in this class is to remove 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 effect on the deghosting quality for this class of approaches.

Khan et al. [62] 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:

$$P(E_n(p)|F) = \sum_{i=1}^{j} a_i \cdot K_{\mathbf{H}}(E_n(p) - \mathbf{y}_i) \Big/ \sum_{i=1}^{j} a_i, \qquad (2.2)$$

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

$$K_{\mathbf{H}}(\mathbf{x}) = |\mathbf{H}|^{-\frac{1}{2}} (2\pi)^{-\frac{d}{2}} \exp(-\frac{1}{2}\mathbf{x}^T \mathbf{H}^{-1}\mathbf{x}), \qquad (2.3)$$

where $\mathbf{H}$ is the bandwidth matrix. In the calculation of Equation 2.2, $\mathbf{y}_i$ belongs to 3×3 neighborhood of the pixel p in *all* exposures. a is initially set to a hat function then both a_p and $P(E_n(p)|F)$ are updated iteratively until convergence:

$$a_{p,t+1} = a_{p,0} \cdot a_{p,t}, \qquad (2.4)$$

where t is the iteration number. At the end of each iteration, $a_{p,t}$ takes the new value of $P(E_n(p)|F)$. After the convergence, the probabilities $P(E_n(p)|F)$ are used instead of the weighting function w in Equation 1.3 to construct the HDR image.

13

Pedone and Heikkilä [100] improve the method of Khan et al. [62] by estimating the parameters of the bandwidth matrix $\mathbf{H}$. Instead of setting it to the identity matrix, the i-th diagonal entry of $\mathbf{H}$ is calculated as:

$$\tilde{h}_i = n^{-\frac{1}{d+4}} \tilde{\sigma}_i, \tag{2.5}$$

where $\tilde{\sigma}$ is the weighted standard deviation of the ith dimension. In addition, due to similarity of some objects with the background, the authors state that using Khan et al.'s algorithm [62], 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. [36] 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 minimizes:

$$\mathcal{E}(\mathcal{L}) = \sum_p D_p(\mathcal{L}_p) + \sum_{(p,q) \in \mathcal{N}} V_{p,q}(\mathcal{L}_p, \mathcal{L}_q) + \sum_{(p,q) \in \mathcal{N}} H_{p,q}(\mathcal{L}_p, \mathcal{L}_q), \tag{2.6}$$

where $D_p(\mathcal{L}_p)$ is the data term, $V_{p,q}(\mathcal{L}_p, \mathcal{L}_q)$ is the smoothness 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 differences. The hard constraint $H_{p,q}(\mathcal{L}_p, \mathcal{L}_q)$ prevents half-included objects by allowing only previously observed labeling transitions. The energy function in Equation 2.6 is minimized via graph cuts [13]. Resulting labeling determines the input source image for each pixel in HDR construction.

Sidibe et al. [120] 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) [23] 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 information to identify the pixels with motion. During HDR image construction phase, only the images in the set S are used.

Silk and Lang [121] introduce a method which employs two different strategies depending on the type of the motion. The algorithm starts with performing a change detection which consists of applying a fixed threshold on the absolute difference of irradiance values in each color channel. This initial motion mask does not respect the object boundaries. In order to refine the initial mask according to the object boundaries, the images are first over-segmented using SLIC superpixels [1] 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 2.3.1.

Zhang and Cham [149, 150] propose an exposure fusion 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 collocated pixels in other exposures.

2.3 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.

2.3.1 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. [61] 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:

$$L_2(p)/L_1(p) = 4, \ \Delta t_2 = 4\Delta t_1. \tag{2.7}$$

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 I is then obtained by fusing the input images as follows:

$$I(p) = \begin{cases} \tilde{L}_1(p), & \text{if } p \text{ is marked or } L_2(p) \text{ is ill-exposed} \\ L_2(p), & \text{otherwise.} \end{cases} \tag{2.8}$$

Grosch [38] provides an extension to Median Threshold Bitmaps (MTB) [140] 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 [39] 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:

$$\left| f\left(\frac{\Delta t_j}{\Delta t_i} f^{-1}(L_i(p)) \right) - L_j(p) \right| < \epsilon \tag{2.9}$$

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. [57] consists of two steps. First, similar to [38], a global image alignment is performed based using MTBs [140]. 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 [71] 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

17

ghost region caused by a stereo mismatch.

Reinhard et al. [108] 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 [99] propose a motion-region detection approach called Bitmap Movement Detection (BMD) which is also based on MTBs [140]. 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:

$$\sum_{i=1}^{N} M_i(p) \notin \{0, N\} \tag{2.10}$$

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 [116], and finally, the detected motion regions are filled from the best-exposed input image for each individual motion region.

Lee et al. [68] propose a histogram-based deghosting method which improves the studies of Min et al. [88, 89] 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:

$$\tilde{r}_n(p) = \texttt{round}\left(\frac{r_n(p) - 1}{R_n - 1} \times 2^B\right), \ 0 \le \tilde{r}_n(p) \le 2^B - 1, \tag{2.11}$$

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. [87] so that only L_{ref} is used to produce output pixels in motion regions.

The algorithm of Silk and Lang [121] was introduced in Section 2.2 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.

2.3.2 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. [32] 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:

$$\ln(L_n(i,j)) = \ln(L_m(i,j)) + \ln(e_{mn}), \tag{2.12}$$

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 [28].

Min et al. [88] 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. [105] 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. [32], 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 [87].

Raman and Chaudhuri [103] improve the simple heuristic of Raman et al. [105] by using a weighted variance measure based on Reinhard et al. [108] and Jacobs et al. [57] 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. [87].

Li et al. [70] 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. [142] 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:

$$\Delta t_i > \Delta t_j \implies L_i(p) \geq L_j(p), \tag{2.13}$$

and the pixel error criterion, which is given by Equation 2.9 in Grosch [38]. 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. [109].

Heo et al. [47] 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 [98, 111] to joint histograms between each pair:

$$\{\langle L_{ref}, L_n \rangle | L_n \in \{L_1, ..., L_N\} - \{L_{ref}\}\}. \tag{2.14}$$

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 [13]. In the next phase, the CRF is estimated using the pixels in the static regions. The HDR reconstruction weights $w_n(p)$ are based on bilateral filtering weights [132] 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. [4] propose another LDR exposure-fusion algorithm. Different from Raman et al. [105] and Raman and Chaudhuri [103], the motion detection operation is embedded into the exposure fusion equation with the following weighting formula:

$$w_n(p) = W_n(p) \cdot Z_n(p) \cdot O_n(p), \tag{2.15}$$

where $W_n(p)$ is the weighting term used by Mertens et al. [87] 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 [104] propose some improvements to their previous study [103] by replacing the exposure fusion approach of Mertens et al. [87] 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. [89] improve their previous motion detection algorithm based on multi-level threshold maps [88] 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 [25] is modified by incorporating a down-weighting term for ghost regions and filtering for noise reduction using a structure adaptive anisotropic filter [77, 78, 143].

Moon et al. [92] handle the ghosting problem by introducing an additional term to the weighting formula of Mertens et al. [87], similar to An et al. [4]. 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:

$$M_n(p) = \exp\left(-\frac{\left(L_{ref}(p) - \tilde{L}_n(p)\right)^2}{2c\sigma_{noise}^2}\right), \qquad (2.16)$$

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. [87] to obtain the enhanced output LDR.

Srikantha et al. [127] 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 [151] improve their previous gradient-based deghosting method [149, 150] 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. [94] 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:

$$
\begin{aligned}
I_n \circ h &= f(k(R + S_n) \cdot \Delta t_i) \\
&= kR \cdot \Delta t_i + kS_n \cdot \Delta t_i \\
&= A_n + S_n,
\end{aligned}
\tag{2.17}
$$

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

23

decomposition is used to recover the artifact-free observation of the scene.

Sung et al. [129] apply a local thresholding to the zero-mean normalized cross correlation (ZNCC) [134] 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 [133]. 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. [35] 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:

$$
\begin{aligned}
\mathcal{E}(F) = \sum_{p} & \left(\mathbf{1}_{\{Pr(p|F(p)) \, < \, \alpha\}} + \gamma \cdot V(F(p))) \right) \\
& + \beta \cdot \sum_{(p,q)\in\mathcal{N}} \mathbf{1}_{\{Pr(p|F(p,q) \, < \, \alpha \, \vee \, Pr(q|F(p,q)) \, < \, \alpha\}},
\end{aligned}
\tag{2.18}
$$

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 [137] normalize the brightness level of all input images to the brightness

24

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:

$$T_n(p) = |\Delta\bar{l}_n|^\beta + \left(\frac{L_n(p) - 50}{15}\right)^2,\qquad(2.19)$$

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. [87] to produce the output.

Lee et al. [67] propose another rank minimization approach which is very similar to Oh et al. [94]. The optimization function used in Lee et al. [67] does not contain homography mappings but instead includes a separate variable as the ghost mask.

2.4 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.

2.4.1 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 [11] introduces a pattern-selective fusion process which uses Laplacian pyramid [15] 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 [49] 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. [147], minimizing the forward and backward flow residuals r_n and r_m:

$$
\begin{aligned}
r_n(p) &= L_m(p - \mathbf{u}_n(p)) - g_{nm}(L_n(p)), \\
r_m(p) &= L_n(p + \mathbf{u}_n(p)) - g_{mn}(L_m(p)).
\end{aligned}
\tag{2.20}
$$

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:

$$
w_n(p) = 0.5 - \tan^{-1}((|r_n(p)| - \mu)/\pi\sigma),
\tag{2.21}
$$

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. [155] 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:

$$\mathcal{E}(\mathbf{u}_n) = \sum_p \left[D(\mathbf{u}_n) + \gamma S(\nabla \mathbf{u}_n) \right], \qquad (2.22)$$

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. [29] 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:

$$|M_n(p) - \mu| > \beta \sigma, \qquad (2.23)$$

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 [58] 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:

$$(\mathbf{d}^*, \mathbf{o}^*, \mathbf{s}^*) = \underset{\mathbf{d}, \mathbf{o}, \mathbf{s}}{\mathrm{argmin}} \{ U(L_i|\mathbf{d}, \mathbf{o}, \mathbf{s}, L_j) + U(\mathbf{d}|\mathbf{o}, \mathbf{s}, L_j) $$
$$+ U(\mathbf{o}|\mathbf{s}, L_j) + U(\mathbf{s}|L_j) \}. \qquad (2.24)$$

$U(L_i|\mathbf{d}, \mathbf{o}, \mathbf{s}, L_j)$ 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. [45] 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. [155]. The energy function is defined as:

$$\mathcal{E}(I, \mathbf{W}) = \int_{\Omega} \left(\sum_{i=1}^{N} M_i + w \cdot \sum_{i=1}^{N} S_{\mathbf{u}_i} + \beta \cdot S_I \right) dx, \qquad (2.25)$$

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_I is the spatial smoothness term of the irradiance map. A coarse-to-fine pyramid structure is used in the minimization to avoid local minima.

2.4.2 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 [86] 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 1.3. Possible mismatches and ghost regions are detected using cross correlation. The weights, $w_n(p)$, of the pixels in these regions are set to zero.

The algorithm of Park et al. [97] 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. [146] 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. [152] 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. [70] (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 [25].

Hu et al. [50] 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. [41]. Then, the IMF is estimated for each color channel by fitting cubic Hermite splines using RANSAC [30] 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 synthesized 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. [87].

Sen et al. [115] propose a PatchMatch [8] 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:

$$
\begin{aligned}
\mathcal{E}(I) = \sum_{p \in pixels} \big[& w_{ref}(p) \cdot (E_{ref}(p) - I(p))^2 \\
& + (1 - w_{ref}(p)) \cdot \mathcal{E}_{MBDS}(I|L_1, ..., L_N) \big],
\end{aligned}
\tag{2.26}
$$

where $\mathcal{E}_{MBDS}$ is the Multisource Bidirectional Similarity (MBDS) measure which is an extension to BDS introduced by Simakov et al. [122]. In case of poorly exposed pixels, $w_{ref}(p)$ 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

30

HDR image directly, auxiliary images are obtained using the search and vote scheme of Simakov et al. [122]. 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. [95] 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 [99], Jacobs et al. [57], Sidibe et al. [120] and Grossberg et al. [39] 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 [51] proposes another PatchMatch [9] 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 [39], the latent images are obtained by minimizing the following energy function:

$$\mathcal{E}(\mathbb{T}, \tau, \mathbf{u}) = C_r(\mathbb{T}, L_{ref}, \tau) + C_t(\mathbb{L}, \mathbb{T}, \mathbf{u}), \quad (2.27)$$

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. [115], Hu et al. does not require the CRFs of the input images to be linear. In one comparison study [135], 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. [153, 154] formulate the patching operation as an optimization problem minimizing the following function for each input image L_n:

$$\sum_{p \in M_n} \| \nabla \Lambda_{n,r(n)}(L_n(p)) - \nabla L_{r(n)}(p) \|_2, \qquad (2.28)$$

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.

2.5   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. [60] 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 [11], they apply a global affine transformation followed by a local optical-flow based correction. For the motion estimation, a variant of the Lucas-Kanade [76] is used, which works on the Laplacian pyramid representation as proposed by Bergen et al. [10]. 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 [113] 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 [46] 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 [79] 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. [60]. The motion estimation process begins with normalizing short exposure frame L_s using:

$$\tilde{L}_l = g^{-1}(g(L_s) - \ln \Delta t_s + \ln \Delta t_l), \tag{2.29}$$

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. [60] 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. [60], 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. [84]. 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 [80].

Castro et al. [18] 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 [140]. 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. [20] provide an application of the exposure fusion [87] 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. [59] introduce a patch-based HDR synthesis method. Similar to Kang et al. [60] and Mangiat and Gibson [79], 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. [115] 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.

CHAPTER 3

SUBJECTIVE EVALUATION

Each class of methods introduced in Chapter 2 has different key characteristics. In order to find out the ranking between a selected subset of these algorithms, the perceived visual quality of their outputs are compared with a subjective experiment. In this chapter, the details of the subjective experiment, including the methodology, benchmarking dataset and the results together with their statistical significance analysis is given.

3.1 The Experiment

The subjective evaluation methodology is similar to the approach followed by Rubinstein et al. [112] where the authors conducted a *pairwise comparison* experiment via a web-based interface. The set of evaluated deghosting algorithms were: (**A**) Grosch [38], (**B**) Khan et al. [62], (**C**) Sen et al. [115], (**D**) Silk and Lang [121], (**E**) Hu et al. [51], (**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 experi-

ment.

The implementations of **A**, **C**, **D**, and **E** were made available by the authors of the algorithms (for **A** as an executable kindly provided 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 [52], it is reimplemented 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 in MATLAB.

3.1.1 Simple Deghosting Algorithm

The simple deghosting algorithm is implementation of a simple ghost detection followed by a Laplacian pyramid [15] 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:

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

(3.1)

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 I is obtained by using the method of Debevec and Malik [25] 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 [15] 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 [140] could be used to align the exposures in a pre-processing step.

(a) Cafe

(b) Candles

(c) FastCars

(d) Flag
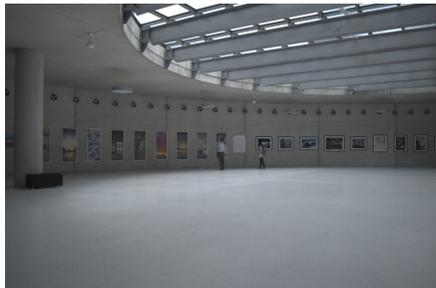
(e) Gallery1
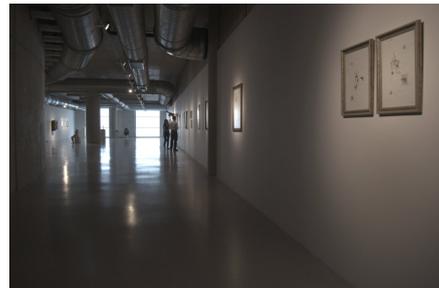
(f) Gallery2

(g) LibrarySide

(h) Shop1

(i) Shop2
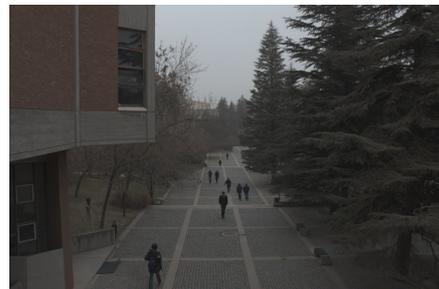
(j) WalkingPeople

Figure 3.1: A representative image for each scene

3.1.2 Benchmark Dataset

For the experiments, 10 different scenes are acquired with different characteristics. The acquisition settings and the image properties of the scenes used in the experiment are given in Table 3.1. 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 [22] software. The output of HDR images were tone mapped using the photographic tone mapping operator [107] for visualization purposes. A representative image for each of these scenes is shown in Figure 3.1.

The priority is given to covering 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.

3.1.3 Experimental Setup

As mentioned above, the experimental design was pairwise comparisons similar to Rubinstein et al. [112]. 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

Table 3.1: 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

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 3.2).

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
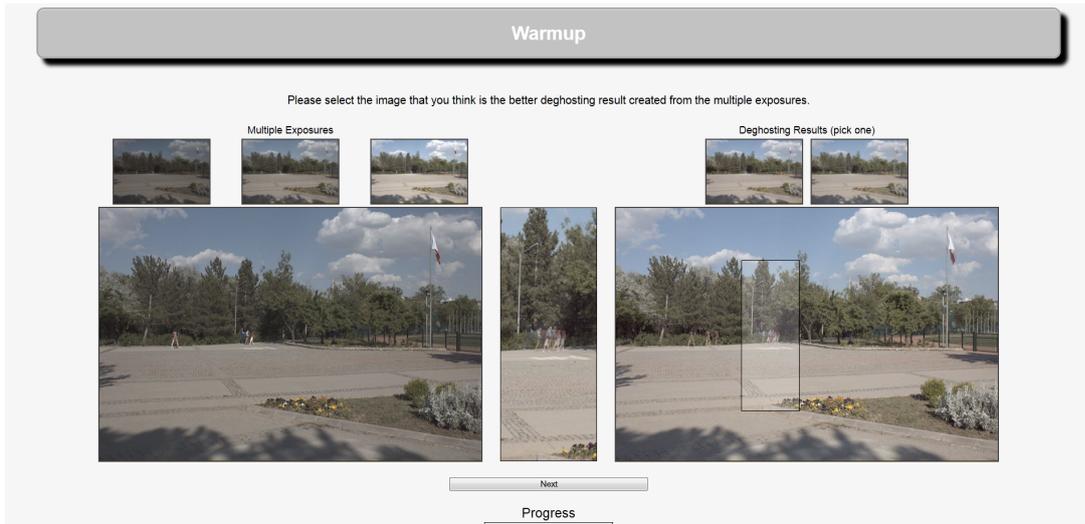
Figure 3.2: Screenshot of the pairwise comparison task.

processing. In total, 63 participants (13F 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.

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, the information about how many times a pair was compared is saved. 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.

3.1.4 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

Table 3.2: 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

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

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

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

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

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

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

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

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

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

a_{ij} is the number of times item i is 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:

$$\pi_i = \frac{1}{t-1} \sum_{j=1, j \neq i}^{t} \pi_{ij}. \tag{3.2}$$

An estimate of π_i is given by:

$$\pi_i = \frac{a_i}{n(t-1)}, \tag{3.3}$$

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 scores, the statistical data analysis method of Starks and David [128] is used which tests the following null hypothesis:

$$H_0 : \pi_i = \pi_j, \forall i, j. \tag{3.4}$$

A special case of the test given by Durbin [26] 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:

$$
\begin{aligned}
D = \sum_{i=1}^{t} d_i^2 &= \frac{4}{nt} \sum_{i=1}^{t} (a_i - \bar{a})^2 \\
&= \frac{4}{nt} \sum_{i=1}^{t} a_i^2 - \frac{1}{4} t n^2 (t-1)^2.
\end{aligned}
\tag{3.5}
$$

Using the formula of Durbin, if D is greater than the critical value χ^2_α for a selected significance level α, it is possible to reject H_0. For a significance value of $\alpha = 0.05$, the corresponding $\chi^2_{0.05} = 12.592$. In the experiment, it was possible to reject H_0 for each scene with D values greater than 100 (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 [128] calculate the smallest amount of statistically significant difference required as:

$$m_c = \lceil 1.96(0.5nt)^{0.5} + 0.5 \rceil. \tag{3.6}$$

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

Table 3.3: 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

of $\alpha = 0.05$. In this study, it is 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 3.2 for each scene and in Table 3.4 for total aggregate preference matrix.

3.2 Results

In this section, first the results of the subjective experiment are presented and interpreted based on the outputs generated by the algorithms. Next, the runtime performance of the algorithms is discussed.

3.2.1 Experimental Results

The results of the subjective experiment for each scene is given in Table 3.2. 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 3.3. The statistical similarity groups for the aggregate results are separately presented in Table 3.4 for clarity purposes.

From these results it is possible to make the following observations. Sen et al.'s [115] and Hu et al.'s [51] methods are clear winners based on the aggregate rankings. These methods also outperform all other methods for each scene with a few exceptions. In

43

Table 3.4: Ranks and significant groups of the algorithms according to the scores (a_i)

Group	Algorithms
1	C - Sen et al. [115] (843), E - Hu et al. [51] (812)
2	F - SimpleDG (759)
3	A - Grosch [38] (656)
4	B - Khan et al. [62] (312)
5	G - NoDG (210), D - Silk and Lang [121] (188)

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 [38] and the simple deghosting algorithm discussed in Section 3.1.1. 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.

An interesting observation is that the simple deghosting algorithm (**F**) explained in Section 3.1.1 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 Walking People 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 [38] 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 3.3 illustrates some of the artifacts created by this algorithm and discusses their causes.

Khan et al.'s method [62] 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 3.4.

The worst performing deghosting algorithm was found to be Silk and Lang's method [121]. Overall, this algorithm was preferred the fewest number of times in pairwise comparisons, receiving a score even lower than the no deghosting condition. This is attributed 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). It is found that these type of artifacts are caused by mathematical singularities during weight estimation in poorly exposed regions. 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 3.5.

In overall, the obtained rankings give confidence about the reliability of the subjec-

Table 3.5: 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

tive 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 [121] algorithm occupied the last position, a finding that is expected from the artifacts in their outputs. Simple deghosting and Grosch's methods [38] 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.

3.2.2 Runtime Performance

Running times of each algorithm is provided in Table 3.5. 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. [115], the "normal" quality setting was used. For Khan et al. [62], the iteration count was set to 10. However, the running times per-iteration are reported as it is observed that $4-5$ iterations were sufficient for convergence, most of the time. All algorithm implementations were in MATLAB, excluding Grosch [38] which was kindly provided to by the author in executable format. The fast running times can be attributed to its being native code and the algorithm's utilization of the GPU.

From Table 3.5 it is possible to 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).
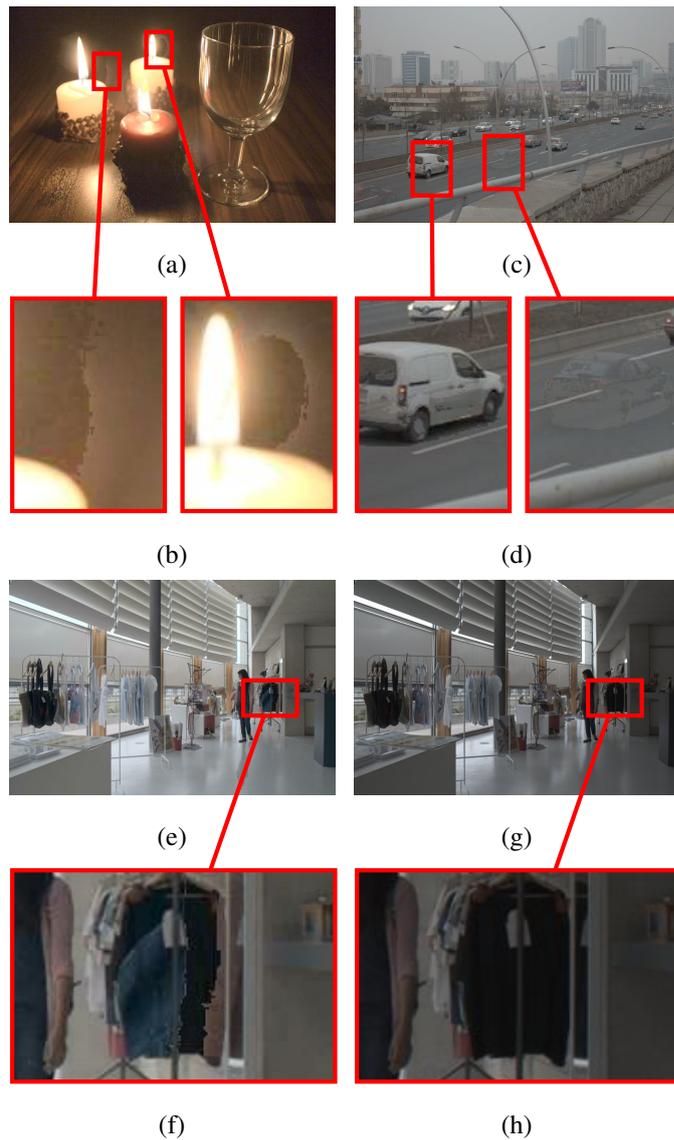
Figure 3.3: Outputs of Grosch [38] for selected scenes. (a), (c), and (e) show the results for Candles, FastCars and Shop2 scenes, (g) shows one of the input LDR exposures for Shop2, (b), (d), (f) and (h) magnify the problematic regions. In (b), the sharp transition between the source images is easily 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.
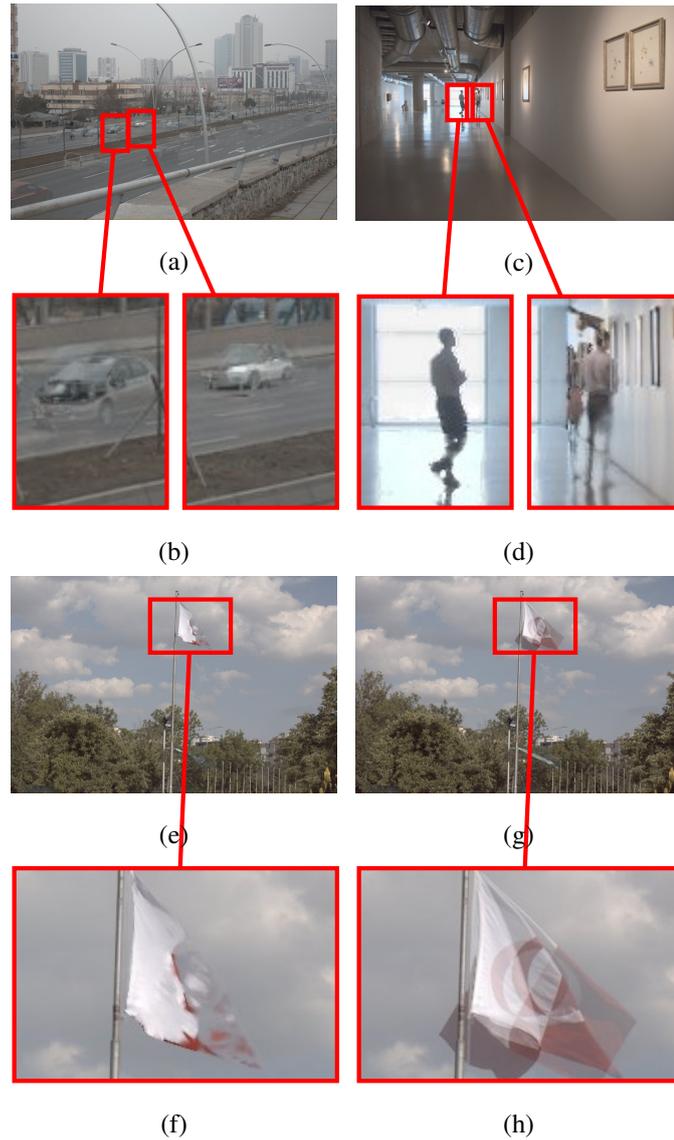
Figure 3.4: Outputs of Khan et al. [62] in (a) FastCars, (c) Gallery2 and (e) Flag scenes. The problematic regions are magnified in (b), (d), (f) and (h). In FastCars scene, the critical assumption of the algorithm does not hold. A vehicle in the scene takes the position of another vehicle from the previous 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.

49
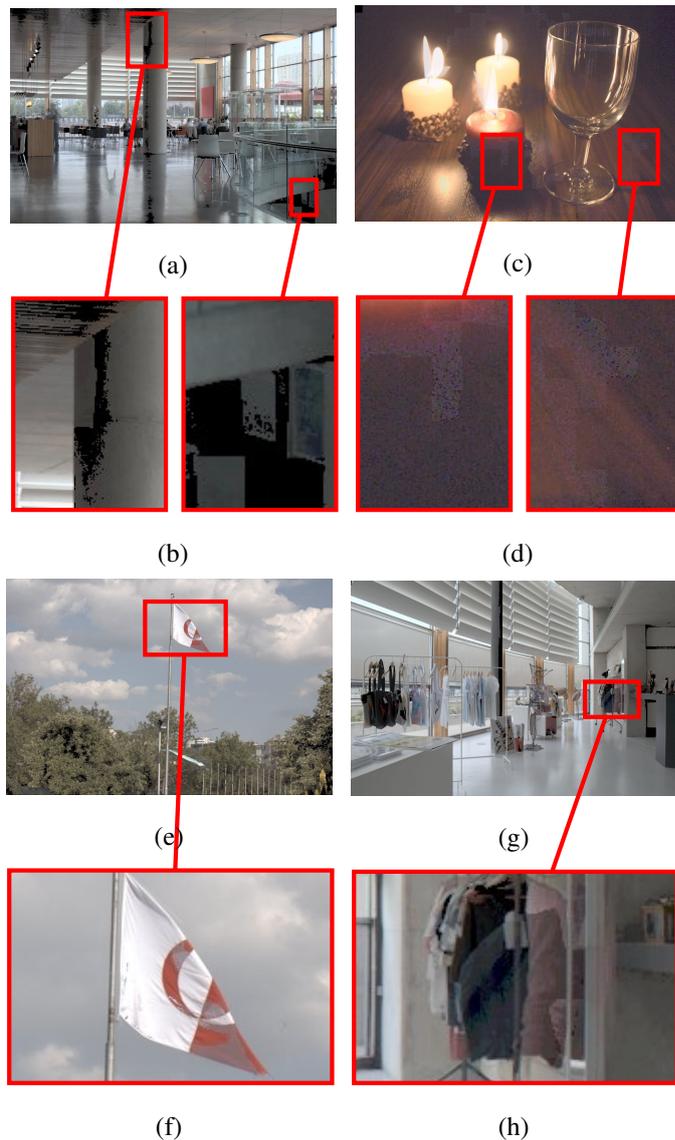
Figure 3.5: Outputs of Silk and Lang [121] for (a) Cafe, (c) Candles, (e) Flag, and (g) Shop2 scenes. In the outputs of the algorithm some of the regions become black, especially if the they have low pixel-intensity values in the input images. In (b), two such regions are shown. In (d), two regions are 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.

CHAPTER 4

OBJECTIVE METRIC AND 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.

The objective metric proposed in this dissertation is the result of analyzing the outputs of several HDR deghosting algorithms to identify the most prevalent artifacts that are present. To validate the compliance of the objective metric with subjective judgements of real observers, a new subjective experiment that involves 16 scenes of varying characteristics, 10 deghosting algorithms, and 52 participants is conducted. Results of this new experiment and new dataset is used for both validation and estimating the best objective metric weights, while the results and dataset from Chapter 3 is used only for validation for improved reliability. It is found that there is a high degree of correlation between the subjective and objective results.

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

4.1 Related Work

Image quality metrics are generally categorized into three classes, namely full-reference (FR), reduced-reference (RR), and no-reference (NR) metrics. FR metrics require a ground-truth reference image in addition to the image whose quality is to be inspected. VDP [24], PSNR [131], VDM [75], SSIM [139], VIF [117], FSIM [148] are some commonly used FR metrics. RR metrics, on the other hand, do not have a ground-truth image but employ some partial information about the reference [12, 69, 125]. Finally, NR metrics do not require any information about the reference. They commonly employ knowledge about specific types of distortions and detect their 'signatures' in the distorted images [141, Chapter 5] [21, 72, 138]. There are some NR metrics as well, which depend on natural scene statistics derived from artifact-free images [91, 118, 123]. Alternatively, a number of studies learn some objective quality assessment functions from a collection of images and their subjective scores or user-selected artifacts and use them to provide NR metric scores [48, 66, 73, 130, 145].

In the literature, there are some studies for building quality metrics for HDR images [108, Chapter 10] [5, 64, 83, 93, 106, 124]. These metrics operate on a pair of images with arbitrarily different dynamic ranges to which the standard quality assessment metrics cannot be directly applied. The metric in this dissertation differs from these previous works in that, it is the first objective quality assessment metric designed to detect deghosting artifacts in HDR images. It can be considered as an RR metric as it requires the individual exposures in addition to the deghosting output.

4.2 Subjective Experiment

In order to validate the compatibility of the objective metrics with subjective preferences, an additional subjective experiment is conducted involving 10 deghosting algorithms and 16 scenes. The selected algorithms were: Grosch [38], Khan et al. [62], Sen et al. [115], Silk and Lang's [121] fluid-motion (FM) and pairwise-downweighting (PWD) methods, Srikantha et al. [127], Hu et al. [51], Lee et al. [67], a simple deghosting approach based on using a single reference image in all dynamic regions (see Section 3.1.1), and no deghosting as a control condition. These algo-
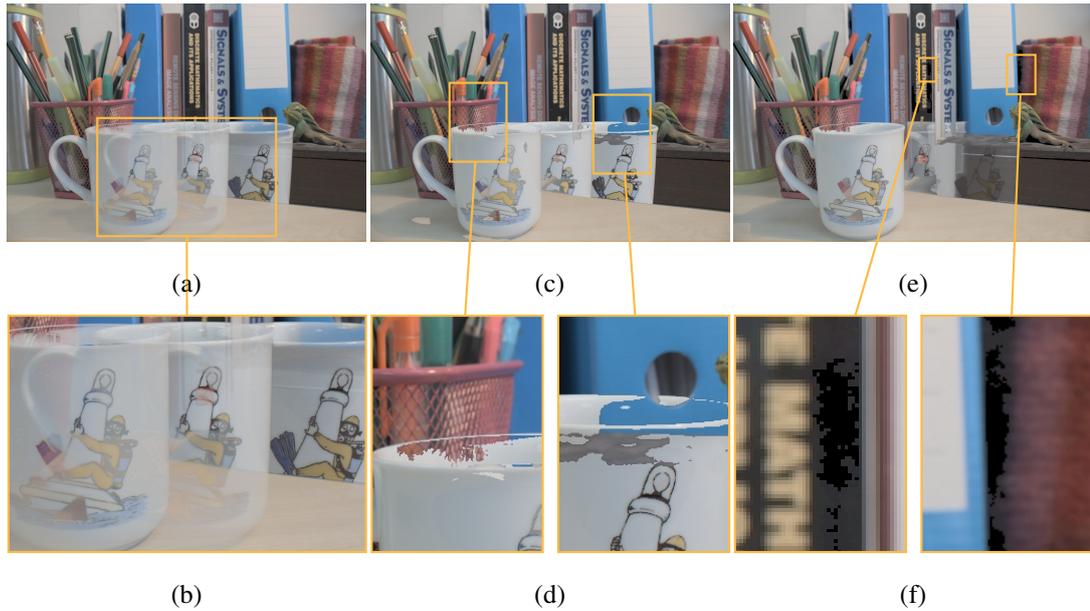
Figure 4.1: Typical deghosting artifacts. (a), (c) and (e) show the blending, gradient inconsistency and visual difference artifacts, respectively with problematic regions magnified in (b), (d) and (f). The images in this figure are obtained by directly merging the exposures without deghosting (a, b), by selecting the best-exposed input image as the only source for each pixel (c, d), and by incorrectly setting the weight values which causes singularities during the HDR assembly process (e, f).

rithms were selected as representatives of different types of deghosting approaches. They range from simple point operations to sophisticated computer vision algorithms.

The source codes and/or executables of each of these algorithms are acquired from the original authors except for Khan et al.'s method [62], was reimplemented in MAT-LAB. In addition, Silk and Lang's algorithm [121] was slightly modified to prevent it producing NaN values due to mathematical singularities in poorly exposed regions.

As for the input scenes, an independent dataset of diverse characteristics is created with both indoor and outdoor environments and motion patterns with varying complexities. Each scene was captured using 3 exposures (-1, 0, +1 EV) with a Canon DSLR camera in RAW format. These scenes are depicted in Figure 4.8 and their properties are summarized in Table 4.3.

Due to a large number of algorithm/scene combinations, an online rating experiment was performed. Similar to the experiment described in Chapter 3, the experiment

started by a warm-up trial to familiarize the participants with the web-interface. The interface consisted of thumbnails of three individual exposures on the left and ten deghosting results on the right. The selected exposure and the deghosting result were shown in higher resolution side-by-side in the middle of the screen. Hovering the mouse over these images brought a zoomed-in view of the region under the cursor to allow detailed analysis.

Different from the subjective experiment described in Chapter 3, instead of performing a pairwise comparison, the participants' task was to give a rating between 0 and 100 for each deghosting result by setting the sliders below each thumbnail. To facilitate these ratings, 0, 25, 50, 75, and 100 values are marked as 'very bad', 'bad', 'medium', 'good', and 'very good' quality, respectively. However, the participants could assign any rating between 0 and 100 with a step size of 5 (e.g. a good image could be rated as 75 whereas a slightly better one as 80). After each trial, the participants viewed the next set of images by pressing the 'Next' button at the bottom of the page. The experiment took between 30 to 40 minutes for each participant. Partially completed experiments are discarded to have equal number of ratings for all conditions. In total, 52 participants finished the experiment in its entirety and their mean ratings are shown in Figure 4.2.

For the statistical analysis of the ratings collected from the subjective experiment, two-way ANOVA for repeated measures is performed with the following factors:

- The deghosting algorithm

- The scene

According to the Mauchly's sphericity test [85], sphericity assumption is violated ($p < 0.005$) and Greenhouse-Geisser [37] correction is applied. The results of the tests of within-subject effects are provided in Table 4.1. It is possible to reject the null hypothesis (average population means being statistically indifferent) with a significance level $p < 0.001$.
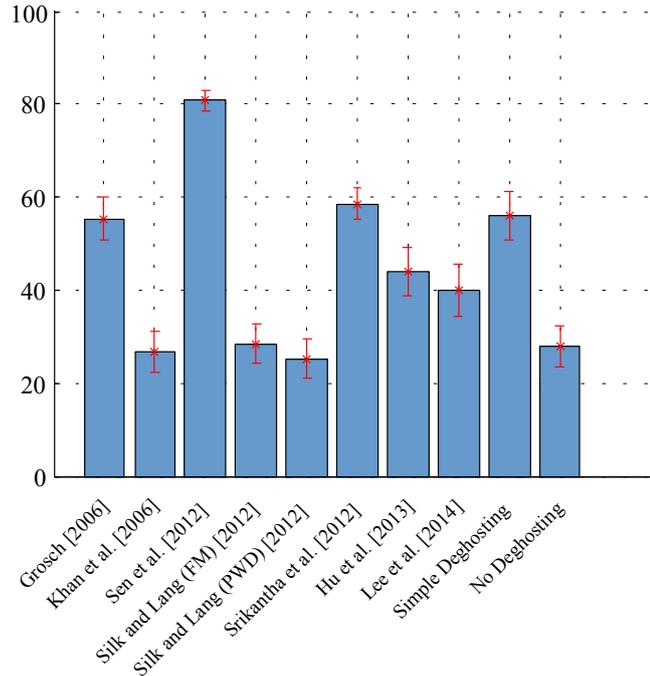
Figure 4.2: Average ratings of the compared algorithms in the subjective experiment. The red lines indicate the standard error.

Table 4.1: Tests of within-subject effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Algorithm	2511773.173	2.827	888645.286	291.624	.000
Error(Algorithm)	439266.490	144.152	3047.235		
Scene	1252215.815	8.093	154734.465	77.394	.000
Error(Scene)	825169.498	412.726	1999.313		
Algorithm * Scene	1331140.724	26.708	49840.727	37.405	.000
Error(Algorithm*Scene)	1814970.214	1362.102	1332.477		

4.3 Deghosting Artifacts

The outputs of several deghosting algorithms for a variety of exposure sequences are studied in order to understand what types of deghosting artifacts are produced. The following four types of artifacts were found to be the most common (Figure 4.1).

Blending. Blending artifacts occur when a dynamic object is blended to its background. This type of artifact is the most common for algorithms which aim to eliminate moving objects, such as Khan et al. [62], but other types of algorithms were also found to exhibit blending artifacts to some extent.

Gradient inconsistency. Gradient inconsistencies occur when the HDR image con-

Table 4.2: Ranks and significance groups of the algorithms according to 95% confidence interval about the average algorithm ratings.

Group	Algorithms
1	Sen et al. [115] (80.91),
2	Srikantha et al. [127] (58.53)
	SimpleDG (56.06)
	Grosch [38] (55.30)
3	Hu et al. [51] (43.80)
	Lee et al. [67] (40.01)
4	Silk and Lang [121]-FM (28.41)
	No deghosting (27.77)
	Khan et al. [62] (26.71)
	Silk and Lang [121]-PWD (25.25)

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

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

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

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

(a) Moving people generate blending (red) and vi-
sual difference (blue) artifacts.

(b) Over-smoothing gives rise to gradient inconsis-
tency (green) artifacts.

Figure 4.3: The objective metric detects several kinds of HDR deghosting artifacts. In
(a), Khan et al.'s [62] output is shown in the bottom-left corner and the metric's result
in the bottom-right. The same for (b), except Hu et al.'s [51] deghosting algorithm
is used. Exposure sequences are shown on the top. Cyan color occurs due to both
gradient and visual difference metrics producing high output.

4.4 Objective Deghosting Metric

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

In all metrics, it is assumed that the input exposures are aligned either by use of a
static camera or by applying a global alignment algorithm such as median threshold
bitmaps [140]. Without this assumption, it would be required to align the images
within the metric, which would make the results dependent on the quality of the
alignment algorithm used.

4.4.1 Blending Metric

Blending artifacts occur when two or more pixels that have different irradiance values
are combined to produce an HDR pixel (Figure 4.1 (a-b)). The presence of this artifact
can be detected if one knows the weights, w, used during the HDR creating process

as these weights determine how much each pixel contributes to the final result:

$$I(p) = k \sum_{n=1}^{N} w_n(p) E_n(p). \tag{4.1}$$

Here, I is the HDR image, p represents the pixel index and k represents a normalization constant[1]. E_n is the nth exposure transferred into irradiance domain by dividing with the exposure time and applying the sensor-specific inverse camera response function.

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

4.4.1.1 Weight Estimation

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

$$\tilde{\mathbf{w}}'(p) = \arg\min_{\alpha} \|I(p) - D(p) \cdot \alpha\|_2^2, \quad \alpha \geq 0, \tag{4.2}$$

where $\tilde{\mathbf{w}}'(p) = [\tilde{w}_1'(p) \, \tilde{w}_2'(p) \, ... \, \tilde{w}_n'(p)]^\mathsf{T}$ is the vector of non-negative weight estimates for pixel p and $D(p)$ is $3 \times N$ dictionary matrix of irradiance vectors from each irradiance map:

$$D(p) = \Big[E_1(p) \mid E_2(p) \mid ... \mid E_N(p) \Big]. \tag{4.3}$$

To comply with Equation 4.1, the weights are normalized to obtain the final weights that will be used in the blending metric: $\tilde{\mathbf{w}}(p) = [\tilde{w}_1(p) \, \tilde{w}_2(p) \, ... \, \tilde{w}_n(p)]^\mathsf{T}$:

$$\tilde{w}_n(p) = \tilde{w}_n'(p) \Big/ \sum_{n=1}^{N} \tilde{w}_n'(p). \tag{4.4}$$

For the least squares estimation, Lawson-Hanson [65] algorithm is used, which recovers the non-negative HDR reconstruction weights via ℓ_1-minimization [31]. This eliminates the need for a regularization term in the weight estimation.

---

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

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

$$w_T(x) = 1 - |2(x - 0.5)|, \tag{4.5}$$

$$w_{BH}(x) = 1 - (2x - 1)^{32}, \tag{4.6}$$

$$w_G(x) = \exp(-25(x - 0.5)^2), \tag{4.7}$$

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

A follow-up experiment is performed to evaluate the similarity of the blending maps obtained using the actual weights and estimated weights. As shown in Figure 4.4, an artificial exposure sequence was created, which contained a single moving object. These exposures are combined using a triangular weighting function and without ghost removal to obtain the results shown in (b). Then the blending map is computed using these *known* weights as shown in (c). Next, the weights are estimated from the HDR image and the input exposures using the weight estimation scheme. The blending map computed from these estimated weights are shown in (d). These two blending maps are visually very similar and have a Pearson correlation coefficient of 0.96 when 3 input exposures are used. Next, 7 input exposures are used to understand how well the weight estimation scheme can deal with a larger number of exposures and a correlation of 0.88 is found between the maps computed from the actual and estimated weights. The HDR image obtained from the estimated weights shown in (e) is also highly correlated with the HDR image reconstructed by using the actual weights (b).

These experiments support that the estimated weights can be reliably used to detect blending artifacts in most cases. The only limitation of the presented weight estima-

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

Another implication of this weight estimation scheme is that if the number of exposures is larger than the number of color channels, the non-negative least squares estimation algorithm performs a sparse recovery of the weights. As a result, the number of non-zero elements in α is limited to 3 in the weight estimation, even if the number of input exposures is greater than 3. However, this does not preclude detecting blending artifacts as shown in the bottom row of Figure 4.4. It simply means that it is possible to detect the blending artifacts caused by the three highest irradiance pixels.

4.4.2 The Metric

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

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

(4.8)

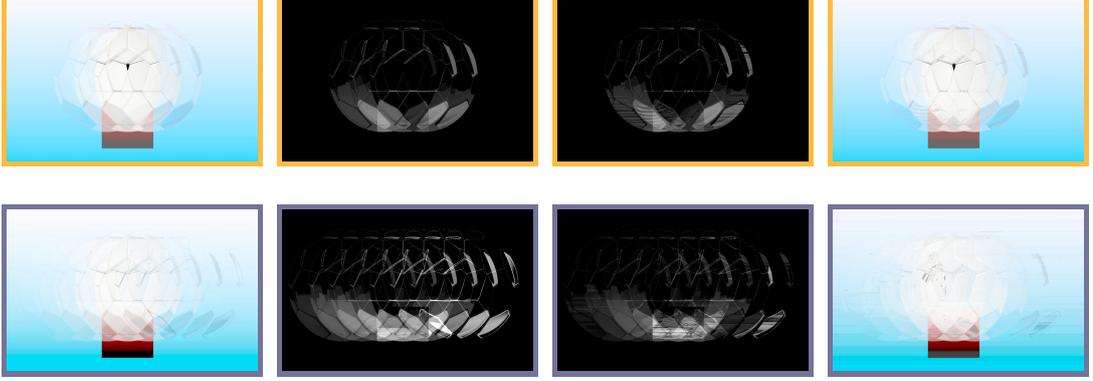where the function g computes the similarity of its inputs:

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

(4.9)

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

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

(4.10)

EV -3   EV -2   EV -1   EV 0   EV +1   EV +2   EV +3

(a) Input Exposures

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

Figure 4.4: Blending metric outputs for 3 and 7 exposures using the actual and estimated HDR reconstruction weights on a set of synthetic input images. It can be observed that the blending metric maps in (c) and (d) are consistent with the visible artifacts in (b).

$$h(E_n(p), E_m(p)) = \frac{\|E_n(p) - E_m(p)\|_2}{\|M_{n,m}(p)\|_2}, \tag{4.11}$$

$$M_{n,m}(p) = \begin{bmatrix} \max\{E_n^r(p), E_m^r(p)\} \\ \max\{E_n^g(p), E_m^g(p)\} \\ \max\{E_n^b(p), E_m^b(p)\} \end{bmatrix}. \tag{4.12}$$

Here, τ represents a tolerance threshold. It is assumed that if the input irradiances are similar, their blending will not cause visible blending artifacts. Various values of τ are experimented and it is found that $\tau = 0.30$ gives the highest correlation with subjective preference. This value is used for all results in this dissertation.

Finally, $W_{n,m}(p)$ represents the joint well-exposedness of a pixel p for exposures n

and m. A broad-hat function is used to represent well-exposedness:

$$w_{BH} = 1 - (2x - 1)^{32}, \tag{4.13}$$

$$W_{n,m}(p) = w_{BH}(L_n(p))w_{BH}(L_m(p)). \tag{4.14}$$

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

4.4.3 Gradient Inconsistency Metric

It is assumed that an HDR image should not contain any gradients that do not exist in any of its constituent exposures. Similarly, if there is a gradient in all exposures, this gradient should exist in the HDR image as well (Figure 4.1 (c-d)). The pixels which fail these requirements are captured in the gradient inconsistency map, which is defined as follows:

$$Q_G(p) = \begin{cases} 0, & \text{if } Q_{G'}(p) \leq \tau \\ Q_{G'}(p), & \text{otherwise.} \end{cases} \tag{4.15}$$

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

$$Q_{G'_{mag}}(p) = \min_n \frac{\left| \frac{\overline{\|\nabla E_n\|_2}}{\overline{\|\nabla I\|_2}} \|\nabla I(p)\|_2 - \|\nabla E_n(p)\|_2 \right|}{\max \left\{ \frac{\overline{\|\nabla E_n\|_2}}{\overline{\|\nabla I\|_2}} \|\nabla I(p)\|_2, \|\nabla E_n(p)\|_2 \right\}}, \tag{4.16}$$

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

As for the gradient orientation, the minimum angle between the directions of gradient vectors is measured:

$$Q_{G'_{dir}}(p) = \min_n \left| \left[(\theta_I(p) - \theta_n(p) + \pi) \bmod 2\pi \right] - \pi \right| / \pi \qquad (4.17)$$

The division by π scales the output value to the range $[0, 1]$. Furthermore, gradient magnitude and orientation inconsistencies are computed in a multi-scale fashion to capture gradient differences in multiple scales. An image pyramid of 5 levels is computed by Gaussian smoothing and downsampling. Equations 4.16 and 4.17 are evaluated for each pixel of each level giving rise to 5 gradient magnitude and orientation maps. Then artifact maps of all levels are upsampled and merged to determine the artifact maps at the finest level. During this process, the inconsistencies in higher levels are given more weight than the inconsistencies in lower levels as they correspond to more important gradient differences. It is found that assigning 4 times more weight to the errors at a higher level than its immediate lower level gives the highest correlation with subjective results. This is expected as ratio of the number of pixels between neighboring pyramid levels is also 4.

4.4.4 Visual Difference Metric

Besides blending and gradient artifacts, deghosting outputs may also contain certain artifacts that may be collectively termed as visual differences. These may be in the form of noise, corruptions, banding, etc. To capture these general types of artifacts, the HDR-VDP-2.2 metric [93] is extended. HDR-VDP-2.2 metric is the latest and most accurate incarnation of the HDR-VDP family, to enable comparisons between a single HDR image and multiple LDR exposures.

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

$$Q_V(p) = \min_n V'(I(p)/\bar{I}, L_n(p)/\overline{L_n}), \qquad (4.18)$$

where V' is the map of the visual difference detection probability generated by the HDR-VDP-2.2. The color encoding parameter of HDR-VDP-2.2 is set to ITU-R

(a) Scene *StreetDay*

(b) Scene *Plants*

(c) Scene *Pendulum*

(d) Scene *ToyTrain*
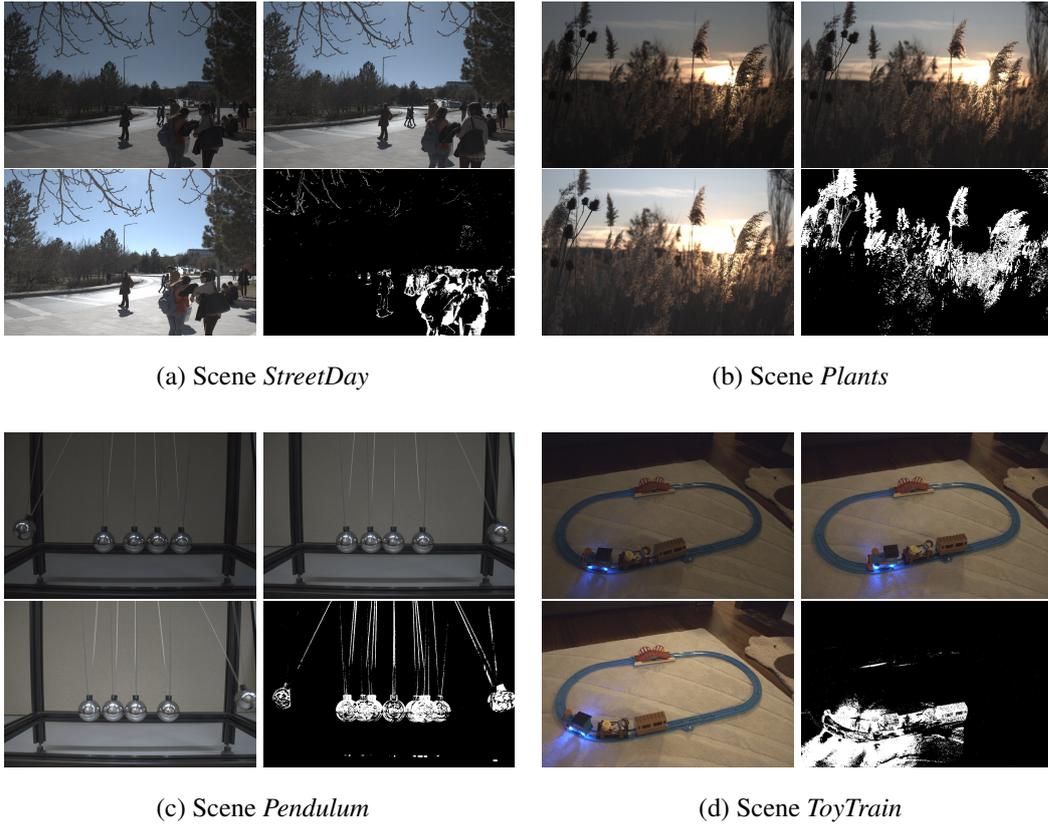
Figure 4.5: Dynamic region bitmaps detected by the heuristic described in the text. Note that the bitmaps successfully capture dynamic objects such as people, foliage, pendulum, and a toy train while generating only a few false positives.

BT.709 RGB and the pixels per degree parameter is given as 30, which is an approximate value for a computer screen with a standard resolution from a typical viewing distance. Note that, this metric is not tailored for a specific type of artifact but it reports errors for any predicted visual differences.

4.4.5   Dynamic Range Metric

The metrics discussed so far do not award an algorithm for producing a higher dynamic range output. However, simply producing a higher dynamic range output is also not sufficient if the output contains visual artifacts. As such, an ideal algorithm should maximize the dynamic range without producing visually disturbing artifacts. In this metric, only the dynamic range in the dynamic image regions is computed. Otherwise, if the static regions have higher dynamic range, they could mask the dy-

(a) Input exposures

(b) Deghosting output   (c) Blending artifact   (d) Grad. magnitude artifact

(e) Grad. direction artifact   (f) Visual difference   (g) Combined visualization
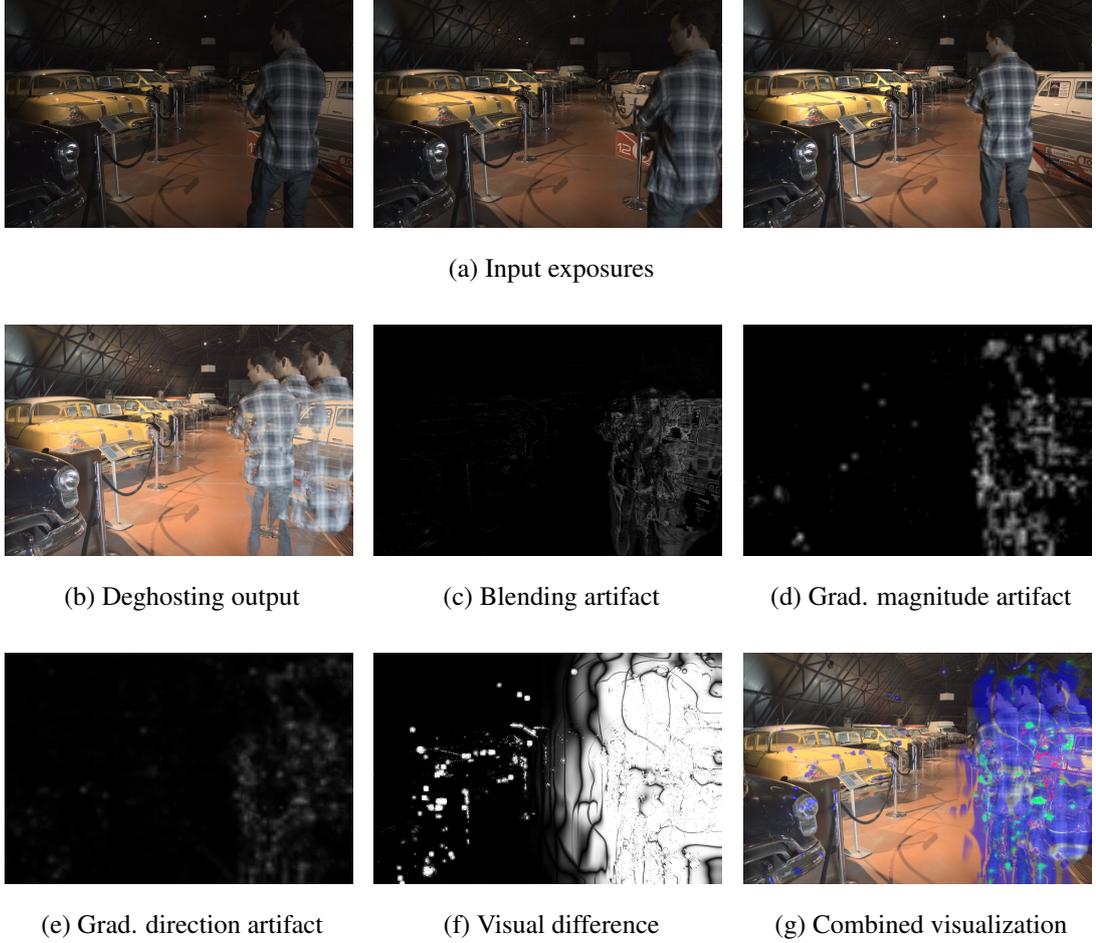
Figure 4.6: Khan et al.'s [62] output showing blending, gradient inconsistency and visual difference artifacts.

namic range in the dynamic regions. To this end, a simple heuristic is used to estimate a dynamic region bitmap:

$$DR(p) = \begin{cases} 0, & \text{if } DR'(p) \leq \tau \\ 1, & \text{otherwise,} \end{cases} \tag{4.19}$$

where

$$DR'(p) = \max_{c \in \{r,g,b\}, n \in \{1,...,N-1\}} h(E_n^c(p), E_{n+1}^c(p))W_{n,n+1}(p) \tag{4.20}$$

In Figure 4.5, the output of this heuristic is demonstrated for several dynamic scenes. The dynamic range, D, is then computed from the HDR image pixels where $DR(p) = 1$. A small percentage of the outliers are excluded to obtain a more stable measure:

$$Q_D = \log_{10} I(p_{99\%}) - \log_{10} I(p_{1\%}). \tag{4.21}$$

(a) Input exposures

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

Figure 4.7: Metric outputs for scene *ToyTrain* using 7 exposures with the algorithm of Silk and Lang [121].

4.5 Results and Validation

In this section, first the visual output of the objective quality metric is demonstrated for several deghosting algorithms applied on different exposure sets. The correlation of the proposed metric is compared with the results of a subjective experiment. Next, it is shown that existing metrics that are not specialized for HDR deghosting are inadequate for detecting deghosting artifacts. Finally, the objective metric is leveraged to illustrate an application called hybrid deghosting in which different algorithms' outputs are merged to obtain a higher quality deghosting result.

4.5.1 Visual Evaluation

Three sample visual outputs of the objective metric are demonstrated in Figures 4.3, 4.6, and 4.7. In Figure 4.3 (a), Khan et al.'s [62] output is shown in the bottom-left corner. The corresponding artifact maps are shown to its right as an overlay on top of the deghosting result. Here, and in all figures in this dissertation, blending artifacts are shown in red, gradient magnitude artifacts in green[2], and visual difference artifacts in blue. It can be seen that the objective metrics detect regions that are affected by deghosting artifacts while generating only a few false positives. In Figure 4.3 (b), the objective metric primarily reports gradient magnitude inconsistencies and visual

---

[2] In overlay visualizations, only gradient direction artifacts are shown for clarity. As will be shown later, they have a significant overlap with gradient magnitude artifacts.
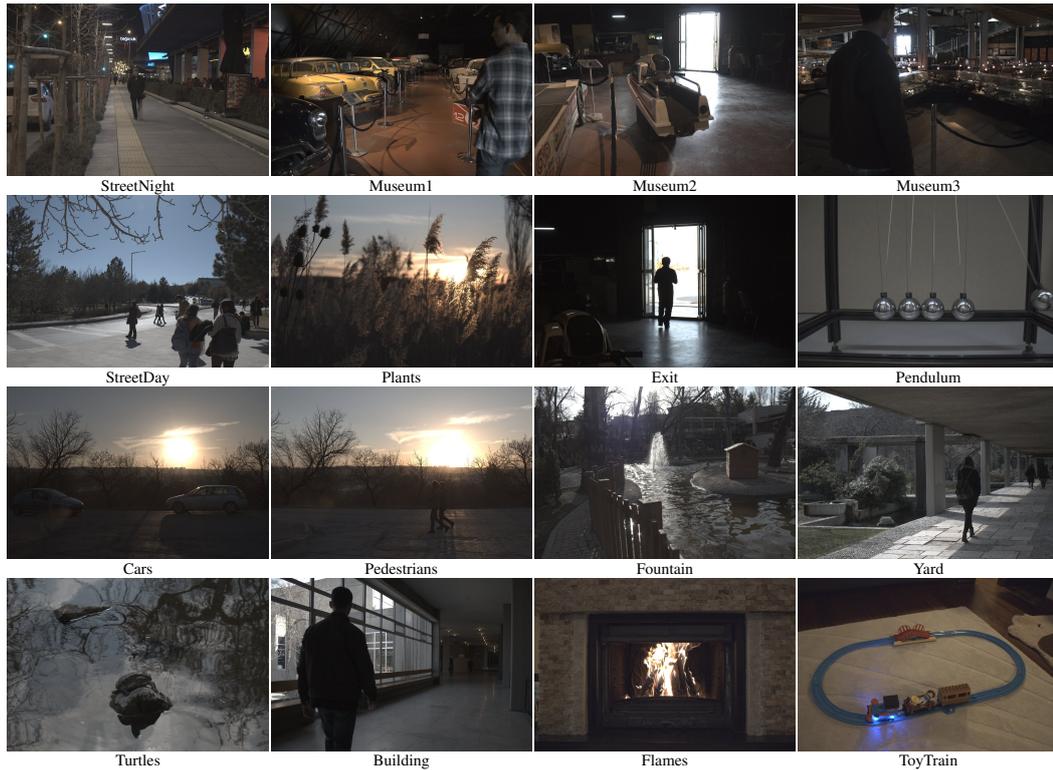
Figure 4.8: The representative images of the input scenes used in the experiment.

differences for the output of Hu et al.'s algorithm [51]. Comparison of this output with individual exposures reveal that there is indeed a loss of details at the back of the person as well as in the distant corridor.

In Figure 4.6, the individual as well as the combined outputs of the objective metric are demonstrated. The input exposures are shown in (a) and the deghosting result in (b). Individual maps are shown in (c) to (f) and the combined result is shown in (g). Again, it is observed that most of the problematic regions in (b) are captured by the objective metric while generating only a few false positives.

Finally, it is shown that the objective metric can be used to detect artifacts in a larger exposure sequence as shown in Figure 4.7. Here, seven input exposures shown in (a) are used to obtain the deghosting result in (b) by Silk and Lang's algorithm [121]. The outputs of the objective metric are shown in (c) to (f).

Table 4.3: Properties of the scenes used in the experiment.

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

4.5.2 Validation

To understand whether these visual observations can be generalized to a larger set of images, a subjective experiment is conducted involving 10 deghosting algorithms belonging to different classes (with one of them being no-deghosting as a control condition), 16 scenes of varying characteristics, and 52 participants. The input scenes are represented in Figure 4.8 and their characteristics are described in Table 4.3. The purpose of this experiment was to collect subjective rating data and then evaluate the correlation of the objective metric outputs with this data. The details and the results of this subjective experiment are given in Section 4.2.

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

$$Q_i = -\sum_{p \in P} Q_i(p)/|P|, \tag{4.22}$$

where each quality metric is substituted for i and $|P|$ represents the number of image pixels. The result is negated to yield more negative scores for worse results. As for

the dynamic range metric, the dynamic range in dynamic regions (Equation 4.21) is directly used. The Pearson and Spearman correlation coefficients are computed between these scores and the aggregate ratings obtained from the subjective experiment.

Furthermore, the metric scores are computed in two modes. In the first mode, the metric outputs are computed as defined by Equation 4.22. In the second mode, first a visual saliency map is computed using Itti et al.'s model [56] using the deghosting output as the input image to this model[3]. Using the saliency maps the visually important image pixels are taken into account while estimating the metric scores. Then the weighted average of the saliency map with the distortion maps is computed:

$$Q_i^s = -\sum_{p \in P} \sqrt{Q_i(p)S(p)}/|P|, \tag{4.23}$$

where $S(p)$ represents the saliency map value. It is found that using saliency only improves correlations for the gradient metrics while having a slightly adverse effect for the blending metric. This could be because the image regions that contain blending artifacts are not that salient due to the softened appearance of objects. The dynamic range metric produces a single scalar value so it is inappropriate to be used in this equation. Finally, the extended VDP metric is a self-contained model which is already based on visual perception, so it is not combined with the saliency map.

As can be seen in Table 4.4 and Table 4.5, all of the objective metrics are positively correlated with the subjective experiment results. In Table 4.4, the aggregate correlations of all metrics except $Q_{G'_{mag}}^s$ are above 0.50 with the extended VDP being the highest (0.68) followed by blending (0.66), gradient direction (0.62), dynamic range (0.59), and gradient magnitude (0.49) metrics. As for the consistency across scenes, blending metric yields the lowest standard deviation (0.14) followed by VDP (0.21), dynamic range (0.21), gradient direction (0.23), and gradient magnitude (0.31). The negative correlation of $Q_{G'_{mag}}^s$ with the participant preferences for scene 'StreetNight' may be attributed to gradient artifacts being perceived as 'correction' of the motion blur from the input exposures by the participants.

Next, the correlation of the individual metrics is computed with each other for all scene-algorithm combinations. This was performed to understand the degree of the

---

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

Table 4.4: Pearson correlations of the individual metrics with subjective ratings.

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

overlap between the outputs produced by different metrics. Table 4.6 summarize these results. According to this table, there is a high degree of correlation between the gradient magnitude and direction metrics. As the latter has a higher mean correlation and lower standard deviation, the gradient magnitude metric is dropped for improved reliability.

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

Table 4.5: Spearman correlations of the individual metrics with subjective ratings.

Scene	Q_B	$Q^s_{G'_{dir}}$	$Q^s_{G'_{mag}}$	Q_D	Q_V
StreetNight	0.70	-0.03	0.04	0.49	0.27
Museum1	0.83	0.85	0.83	0.72	0.60
Museum2	0.45	0.20	0.38	0.73	0.68
Museum3	0.61	0.34	0.21	0.64	0.51
StreetDay	0.83	0.45	0.60	0.70	0.83
Plants	0.52	0.65	0.53	0.77	0.75
Exit	0.64	0.36	0.18	0.61	0.44
Pendulum	0.92	0.76	0.67	0.71	0.84
Cars	0.60	0.20	0.47	0.74	0.46
Pedestrians	0.27	0.44	0.20	0.88	0.86
Fountain	0.71	0.39	0.15	0.50	0.67
Yard	0.63	0.24	0.46	0.36	0.67
Turtles	0.38	0.73	0.82	0.84	0.83
Building	0.49	0.84	0.95	0.55	0.72
Flames	0.14	0.76	0.61	0.47	0.64
ToyTrain	0.89	0.88	0.83	0.76	0.95
Average	0.60	0.50	0.50	0.65	0.67
Std. Dev.	0.22	0.28	0.28	0.15	0.18

individual metrics:

$$Q_U = [Q_B \; Q^s_{G'_{dir}} \; Q_V \; Q_D \; 1] \cdot [w_B \; w_G \; w_V \; w_D \; w_c]^\mathsf{T}, \tag{4.24}$$

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

In order to find the best weights, adaptive simulated annealing [55] is used for maximizing the average Pearson correlation between Q_U and subjective ratings. For simulated annealing, the initial temperature is set to $T_0 = 100$ and the following annealing schedule is used:

$$T = 0.95^k T_0, \tag{4.25}$$

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

$$P = \frac{1}{1 + \exp\left(\dfrac{\Delta E}{\max\limits_i(T_i)}\right)}, \tag{4.26}$$

where ΔE is the difference between the present and past values of the energy function and T_i is the current temperature of component i. No upper and lower bounds are

Table 4.6: Pearson correlation coefficients between absolute metric scores in all scene-algorithm combinations.

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

imposed on the metric weights; however, the resulting value of vector $\mathbf{w}$ is normalized to have a length of 1. The normalization has no effect on the value of the energy function (i.e. average Pearson correlation) and it is only performed to have metric weights with comparable scales in different scenes for analysis.

Leave-one-out cross-validation (LOOCV) is performed to help validate if the weights learned from all images except one is a good indicator for the left-out image. The individual weights estimated by using adaptive simulated annealing after leaving out each one of the images is given in Table 4.7. It is observed that the estimated weights are consistent with each other. The mean LOOCV value is found to be 0.77 with a standard deviation of 0.14. This high correlation and low standard deviation suggest that the weights learned from a subset of the scenes (training) can be used to estimate the quality of outputs that are not part of the dataset (testing). Therefore, to produce the final set of weights all scenes are included in the optimization, which resulted in the weights shown in Table 4.8.

As an additional validation of the proposed weights, they are used to compute a UDQM score for a set of images used in the independent subjective experiment whose details are given in Chapter 3. The goal was to investigate how well the computed scores correlate with the subjective responses from an entirely independent dataset. The rating of an algorithm is computed as the number of times an algorithm was selected over another one, and then correlated the UDQM scores with those ratings. The resulting correlations are given in Table 4.9. As can be seen from this table, a mean Pearson correlation of 0.65 is obtained with 0.20 standard deviation. Although the mean correlation was somewhat smaller than the correlation of the subjective experiment described in this chapter, it is still high enough to be predictive of the

Table 4.7: Best metric weights found in LOOCV and their corresponding Pearson correlations in training and testing scenes. The scene which is left out is given in the first column.

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

algorithms' performance.

Finally, an analysis is performed to understand how well an individual participant's responses correlate with the mean responses of all participants except herself. As the number of participants was 52, this produced 52 correlation scores with the mean score being 0.75 and the standard deviation 0.22. Given that the mean UDQM correlation for the test set was 0.77 with a standard deviation of 0.14, it is not unreasonable to assume that the combined metric's responses resemble that of an average observer. In addition, by observing close correlation scores, it is possible to state that the UDQM is consistent in a significant amount with the human quality perception patterns from the subjective experiments and contribution of additional auxiliary quality metrics would be insignificant since the correlation score of an average observer is reached.

Table 4.8: Best metric weights obtained using adaptive simulated annealing [55].

w_B	w_G	w_D	w_V	w_c
0.427	0.811	0.029	0.037	0.397

Table 4.9: Pearson correlation coefficients of Q_U from the dataset of Tursun et al. [136] using the weights in Table 4.8.

Cafe	Candles	FastCars	Flag	Gallery1	Gallery2
0.90	0.67	0.64	0.27	0.64	0.61

LibrarySide	Shop1	Shop2	WalkingPeople	Average	Std. Dev.
0.95	0.66	0.43	0.71	0.65	0.20

4.5.3 Comparison with Other Quality Metrics

In this section, the outputs of the two of the potentially relevant quality metrics are demonstrated, namely the dynamic range independent quality metric (DRIM) [5] and a blind deblurring quality metric [73]. The former metric is selected as it can compare images with different dynamic ranges and the latter is selected because deblurring artifacts may be somewhat similar to deghosting artifacts.

The distortion map produced by the DRIM for the images shown in Figure 4.3 are shown in Figure 4.10. To compute these results, the middle exposure of the exposure sequence is compared against the deghosting output. In DRIM outputs, red indicates reversal of contrast, green loss of contrast, and blue amplification of contrast. Although the DRIM metric detects problematic regions it also generates many false positives which correspond to mostly static parts of the scenes (blue regions). As such, the DRIM metric is not suitable to be used for detecting deghosting artifacts.

The other metric that is tested was Liu et al.'s [73] 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 blind deblurring metric of Liu et al. [73] produces a single quality score instead of a distortion map. In order to calculate this quality score, all input HDR images are tone mapped with the tone mapping operator of Reinhard et al. [107] and gamma corrected ($\gamma = 2.2$). For no-reference ringing detection, the original metric uses the

74

Table 4.10: Pearson correlations for Liu et al.'s [73] deblurring metric with the subjective experiment.

Exit	Pendulum	Cars	Pedestrians	Fountain	Yard
0.16	-0.01	0.30	0.64	0.11	0.50

Turtles	StreetNight	Building	Flames	ToyTrain	Museum1
0.76	0.08	-0.24	-0.22	0.24	0.22

Museum2	Museum3	StreetDay	Plants	Avg.	Std.
0.36	-0.21	0.33	0.58	0.23	0.30

Table 4.11: The objective scores obtained using Liu et al. [73] (larger is better) and Spearman's rank correlation ρ with the subjective experiment scores. The average correlation $\bar{\rho} = -0.001$.

	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

blurry input image in addition to the deblurred images. As blurry input image does not exist in this scenario, instead the image obtained by directly merging the exposures is used without applying deghosting. The quality scores are generated for all scenes used in the subjective experiment and the correlations with participant ratings are computed. As shown in Table 4.10, this metric has a low correlation with subjective ratings and therefore is not suitable to be used as a deghosting quality metric.

For improved reliability this metric is also tested on the dataset previously introduced in Chapter 3 and results are provided in Table 4.11. In general, it can be seen that there is low correlation, a finding which confirms that Liu et al.'s metric is not suitable for evaluating HDR deghosting algorithms.

4.5.4 Comparison with Ground Truth

In order to show how well the objective metric captures the visual artifacts, ground truth HDR images are prepared and the absolute difference map is created between the deghosting result and one of the ground truth HDR images for comparison. For this purpose, three exposures are captured with EV -1, 0 and +1 for three different positions of an object in Figure 4.9. After this process, it was possible to create a ground truth HDR image for each position of the object in the scene. Then the deghosting algorithm of Silk and Lang [121] is applied to create an HDR image using EV -1, 0 and +1 from object positions 1, 2 and 3, respectively. Ground truth HDR 2 is chosen as the reference for creating the difference map since it is visually the most similar ground truth image. The minimum and 99th percentile absolute radiance differences are scaled between 0-1 for visualization in the difference map. It is observed that the highest responses of the objective metrics correspond to the pixels with largest absolute differences. The visual difference metric is more sensitive to the artifacts over a larger area while blending and gradient inconsistency metrics give more local responses.

4.5.5 Application: Hybrid Deghosting

Finally, it is demonstrated how to leverage the objective metric for combining the outputs of multiple deghosting algorithms to obtain a higher quality deghosting result. To perform hybrid deghosting, the UDQM map is computed for each one of the input HDR images. Then the binary mask of pixels is extracted from each one of the input HDRs, which will contribute to the hybrid HDR. A pixel is marked in the corresponding binary mask if it has the best UDQM value among all the input HDR images. Before combining the images by using Laplacian blending by using these masks, histogram matching is applied to ensure that the irradiance distributions of the two images are similar.

Two sample outputs are provided in Figures 4.11 and 4.12. In Figure 4.11, Lee et al.'s [67] output is combined with the simple deghosting approach (see Section 3.1.1), which involves selecting the middle exposure in all dynamic regions. In Figure 4.12,

on the other hand, Hu et al. [51]'s and Lee et al.'s [67] results are combined using their distortion maps. In both figures, the combined outputs have fewer artifacts than the individual results.

It should be noted, however, that hybrid deghosting is not suitable to be used for all pairs of algorithms, especially for those which select different reference exposures. In such cases, multiple copies of the same object could appear in the combined result. An example of this limitation is shown in Figure 4.13 where the inconsistency of the HDR images due to the difference in object positions result in an hybrid HDR with noticeable blending and gradient inconsistency artifacts.

Pos. 1 (EV -1) Pos. 1 (EV 0) Pos. 1 (EV +1)

Pos. 2 (EV -1) Pos. 2 (EV 0) Pos. 2 (EV +1)

Pos. 3 (EV -1) Pos. 3 (EV 0) Pos. 3 (EV +1)
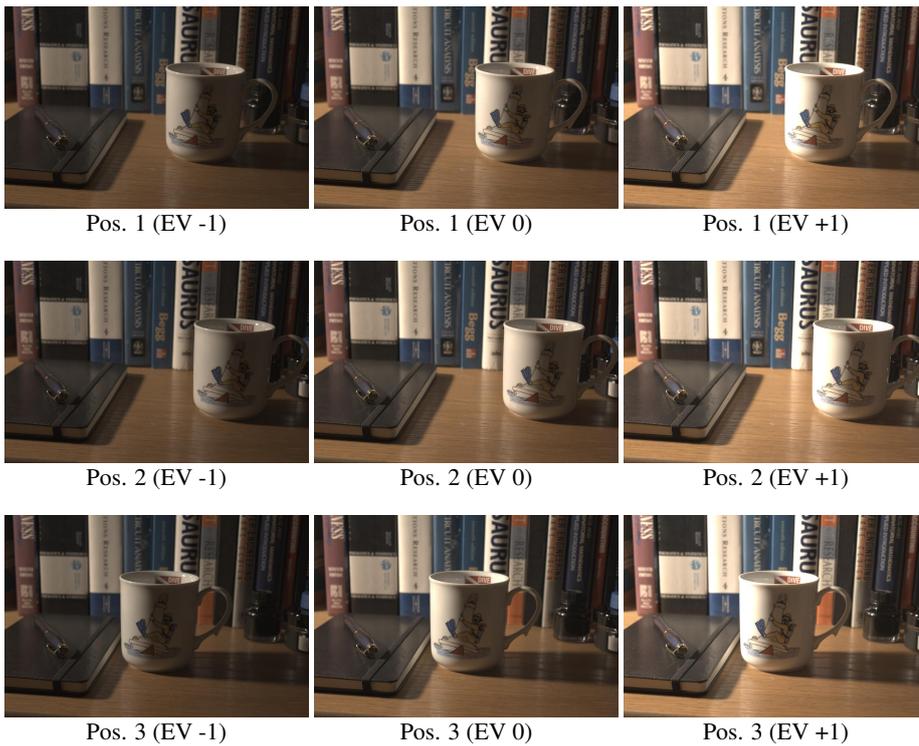
(a) Input exposures for three different positions of the mug

(b) Ground truth 1 (c) Ground truth 2 (d) Ground truth 3

(e) Output HDR (f) Difference Map (g) The metric results
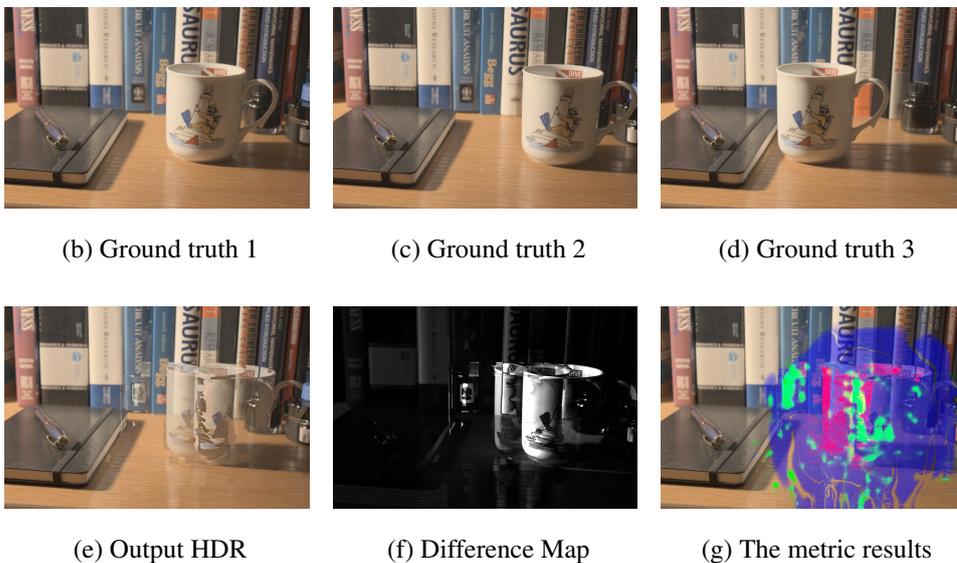
Figure 4.9: Comparison of the objective metric result (g) with the absolute difference (f) between the output of Silk and Lang [121] (e) and ground truth HDR (c). The input exposures with EV -1, 0 and +1 are shown in (a). For each position of the moving object, the ground truth HDR image obtained by using the input exposures is given in (b-d).
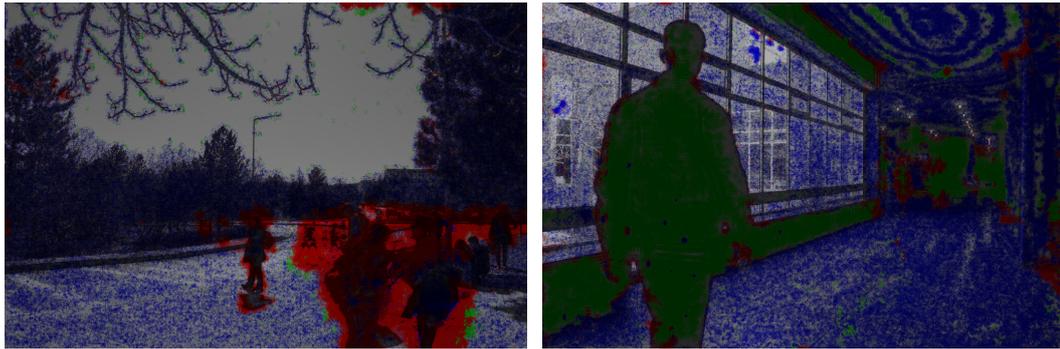
(a) DRIM (Khan et al. [62])   (b) DRIM (Hu et al. [51])

Figure 4.10: The DRIM [5] outputs for two sample scenes. In DRIM, blue represents amplification of contrast, green loss of contrast, and red reversal of contrast. The objective metric results were shown in Figure 4.3.
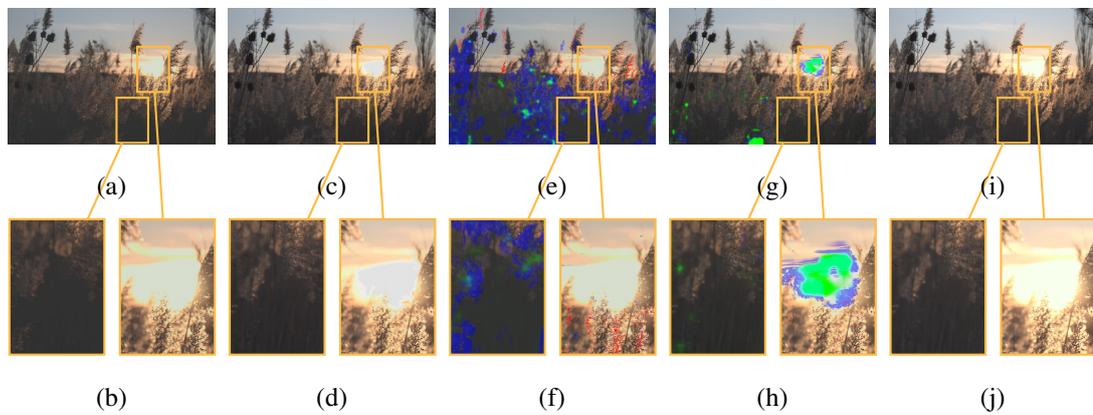
Figure 4.11: The outputs of Lee et al. [67] (a-b) and simple deghosting approach (c-d), the objective metric results for Lee et al. [67] (e-f) and simple deghosting approach, and the hybrid deghosting result (i-j).
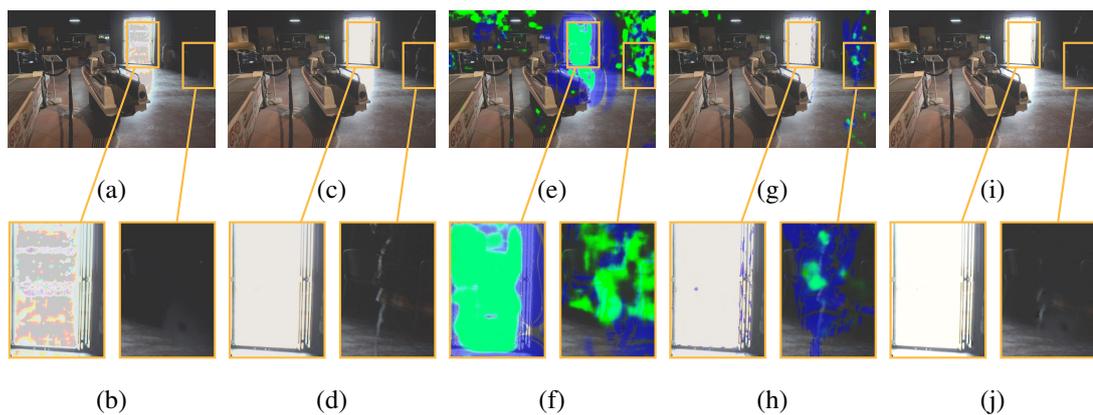
Figure 4.12: The outputs of Hu et al. [51] (a-b) and Lee et al. [67] (c-d), the objective metric results for Hu et al. [51] (e-f) and Lee et al. [67], and the hybrid deghosting result (i-j).

(a)   (b)   (c)
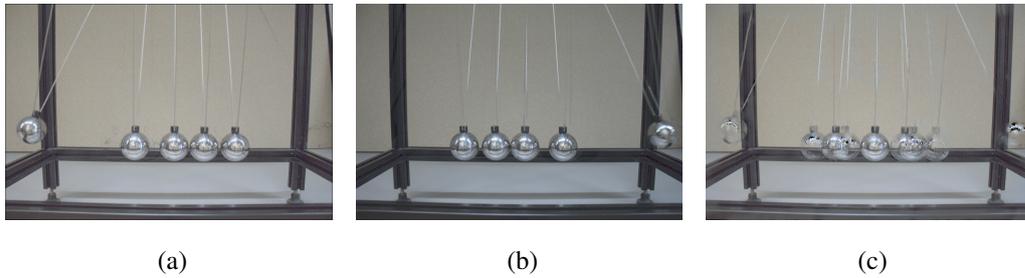
Figure 4.13: The hybrid HDR (c) obtained from two HDR outputs of Srikantha et al. [127] (a) and Lee et al. [67] (b), which are inconsistent with each other in terms of object positions.

CHAPTER 5

DISCUSSION

In both of the subjective experiments, it is observed that the patch-based deghosting algorithms provide the best visual quality. The algorithm of Sen et al. [115] had the first ranking due to its mostly artifact-free results. The main strength of this algorithm may be attributed to its deghosting strategy, which aims at using as much information as possible from the reference input image when it is correctly exposed. One of the weaknesses of this algorithm is that the underlying patch-match algorithm [122] may perform poorly in the over-exposed regions of the reference input. In addition, although performance improvements are introduced by the recent implementations [8, 9], the patch-match operation is still computationally very intensive, making this type of deghosting algorithms unsuitable for the platforms with limited computational power and video deghosting applications.

Simpler algorithms based on moving object selection provide a fast solution to the deghosting problem. However, the number of input images used while reconstructing the dynamic regions limits the amount of dynamic range recovered from the dynamic areas of the scene. In addition, boundary artifacts severely deduce the perceived output quality unless an effective blending method is employed. It is observed that severity of such artifacts highly depends on the scene content.

The poor performance of moving object removal algorithms in the subjective experiments may be related to the small number of input exposures used in the subjective experiments. The performance of algorithms in this class is very sensitive to the number of input exposures capturing the static background. Therefore, they are not a good choice if the number of exposures is limited or if the scene is constantly chang-

ing among the input exposure set. In addition, iterative methods in this class perform pixel-wise weight adjustment operations which may take a long time to converge. Nevertheless, the algorithms in this class are the only option if it is desired to remove the moving objects in the HDR scene and estimate the static background.

In both of the subjective experiments, it is surprising that a simple artifact-free deghosting algorithm such as the simple deghosting method achieved a high rank. With this observation it is possible to state that the participants often put a higher priority on artifact-free outputs over the limited dynamic range. This finding is also confirmed by the recent work of Gryaditskaya et al. [40].

Although within-subject designs is a popular methodology for benchmarking, there are several challenges in the evaluation of the deghosting algorithms. First of all, the definition for the "best" algorithm is highly application-dependent. In some cases observers may place a higher priority on the preservation of the moving objects in the outputs. While in some other cases it is possible to observe that the opposite is desirable (e.g. in architectural photography). Therefore, the ranking of the algorithms is highly dependent on expectations from the deghosting algorithm in specific application areas.

Second, post-processing operations are an important factor which may have an effect on the ranking. In the subjective experiments the global tone mapping operator of Reinhard et al. [107] is used for preserving the original stimuli as much as possible. However, in real-world conditions use of different tone mapping operators may have a large impact on the preferences of the observers by masking or amplifying some types of artifacts. In addition, observing the outputs in an actual HDR display without any post-processing is probably going to change the ranking of the algorithms as the observers will be able to experience the actual dynamic range provided by the deghosting algorithms.

CHAPTER 6

CONCLUSION

In this dissertation, 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, approximately 50 HDR deghosting algorithms are classified. The distinguishing characteristics of the algorithms within each class are provided. The algorithms are also evaluated in two subjective experiments on different benchmark datasets. The benchmark datasets and the proposed evaluation framework are suitable to be used in future evaluation studies as well.

This dissertation aims at fulfilling the growing need for the evaluation of recently proposed methods due to the increasing number of algorithms in this field. To this end, an objective HDR deghosting quality metric, comprised of individual metrics tuned for different types of artifacts is presented and it is shown that they correlate well with visual observations and subjective preferences. The first application that the objective metric enables is automatic quality evaluation of deghosting algorithms. By using the objective metric, one can often avoid comprehensive subjective experiments which are both tedious and may become outdated as new algorithms are proposed. Secondly, the proposed metrics can be used to optimize parameter selection for deghosting algorithms. For instance, a deghosting method of choice may be run in batch-processing mode and allowed to explore a parameter space until it finds the combination that gives the least distortions as computed by the metric. The objective metric can also be used to rapidly assess the strengths and weaknesses of different algorithms for different images sets and ultimately design improved algorithms. Finally, as demonstrated in the last chapter, the objective metric also enables a novel

application called hybrid deghosting, whereby multiple deghosting results are combined to obtain a higher quality result.

Despite the large amount of HDR deghosting algorithms proposed during the last decade, the HDR deghosting is still a popular problem that is being worked on. As the trend is shifting towards mobile computation there is still room for improvement in this field. In addition, HDR video deghosting is still an unsolved problem. The HDR image deghosting algorithms cannot be applied to videos in their current form, since the temporal dimension of the video data must be taken into account. Failing to do so, results in different types of artifacts such as flickering and semantic inconsistencies between subsequent frames of an HDR video.

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

REFERENCES

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. SLIC superpixels. Technical report, EPFL Technical Report, 2010.

[2] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen. Interactive digital photomontage. *ACM Transactions on Graphics*, 23(3):294–302, 2004.

[3] A. O. Akyüz. Photographically guided alignment for HDR images. *Eurographics-Areas Papers*, pages 73–74, 2011.

[4] J. An, S. H. Lee, J. G. Kuk, and N. I. Cho. A multi-exposure image fusion algorithm without ghost effect. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1565–1568. IEEE, 2011.

[5] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H.-P. Seidel. Dynamic range independent image quality assessment. *ACM Transactions on Graphics*, 27(3):69:1–69:10, Aug. 2008.

[6] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1):1–31, 2011.

[7] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers. *Advanced High Dynamic Range Imaging: Theory and Practice*. CRC Press (AK Peters), Natick, MA, first edition edition, 2011.

[8] C. Barnes, E. Shechtman, A. Finkelstein, and D. Goldman. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics*, 28(3):24, 2009.

[9] C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein. The generalized PatchMatch correspondence algorithm. In *European Conference on Computer Vision*, pages 29–43. Springer, 2010.

[10] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *European Conference on Computer Vision*, pages 237–252. Springer, 1992.

[11] L. Bogoni. Extending dynamic range of monochrome and color images through fusion. In *15th International Conference on Pattern Recognition*, volume 3, pages 7–12. IEEE, 2000.

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

[13] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

[14] M. Brown and D. G. Lowe. Recognising panoramas. In *Proceedings of the 9th IEEE International Conference on Computer Vision*, ICCV, Washington, DC, USA, 2003. IEEE Computer Society.

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

[16] P. J. Burt and R. J. Kolczynski. Enhanced image capture through fusion. In *4th International Conference on Computer Vision*, pages 173–182. IEEE, 1993.

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

[18] T. Castro, A. Chapiro, M. Cicconet, and L. Velho. Towards mobile hdr video. *Proceedings of the Eurographics-Areas Papers*, pages 75–76, 2011.

[19] L. Cerman and V. Hlavac. Exposure time estimation for high dynamic range imaging with hand held camera. In *Proceedings of Computer Vision Winter Workshop, Czech Republic*. Citeseer, 2006.

[20] A. Chapiro, M. Cicconet, and L. Velho. Filter based deghosting for exposure fusion video. In *ACM SIGGRAPH Posters*, page 33. ACM, 2011.

[21] C. Chen, W. Chen, and J. A. Bloom. A universal reference-free blurriness measure. *Proceedings of SPIE*, 7867:78670B–78670B–14, 2011.

[22] D. Coffin. DCRaw: Decoding raw digital photos in linux. Accessed: February 2014.

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

[24] S. Daly. The visible differences predictor: An algorithm for the assessment of image fidelity. In A. B. Watson, editor, *Digital Images and Human Vision*, pages 179–206. MIT Press, 1995.

[25] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH, pages 369–378. ACM Press/Addison-Wesley Publishing Co., 1997.

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

[27] A. Eden, M. Uyttendaele, and R. Szeliski. Seamless image stitching of scenes with large motions and exposure differences. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 2498–2505, 2006.

[28] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. *ACM Transactions on Graphics*, 21(3):249–256, 2002.

[29] S. Ferradans, M. Bertalmío, E. Provenzi, and V. Caselles. Generation of HDR images in non-static conditions based on gradient fusion. In *VISAPP*, pages 31–37, 2012.

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

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

[32] O. Gallo, N. Gelfand, W.-C. Chen, M. Tico, and K. Pulli. Artifact-free high dynamic range imaging. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–7. IEEE, 2009.

[33] M. Gevrekci and B. K. Gunturk. On geometric and photometric registration of images. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 1. IEEE, 2007.

[34] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H.-P. Seidel, and H. Lensch. Optimal HDR reconstruction with linear digital cameras. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 215–222. IEEE, 2010.

[35] M. Granados, K. I. Kim, J. Tompkin, and C. Theobalt. Automatic noise modeling for ghost-free HDR reconstruction. *ACM Transactions on Graphics*, 32(6):201, 2013.

[36] M. Granados, H.-P. Seidel, and H. Lensch. Background estimation from non-time sequence images. In *Proceedings of Graphics Interface*, pages 33–40. Canadian Information Processing Society, 2008.

[37] S. W. Greenhouse and S. Geisser. On methods in the analysis of profile data. *Psychometrika*, 24(2):95–112, 1959.

[38] T. Grosch. Fast and robust high dynamic range image generation with camera and object movement. In *Proceedings of Vision Modeling and Visualization*, pages 277–284, 2006.

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

[40] Y. Gryaditskaya, T. Pouli, E. Reinhard, K. Myszkowski, and H.-P. Seidel. Motion aware exposure bracketing for hdr video. In *Computer Graphics Forum*, volume 34, pages 119–130. Wiley Online Library, 2015.

[41] Y. HaCohen, E. Shechtman, D. B. Goldman, and D. Lischinski. Non-rigid dense correspondence with applications for image enhancement. *ACM Transactions on Graphics*, 30(4):70, 2011.

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

[43] K. K. Hadziabdic, J. H. Telalović, and R. Mantiuk. Comparison of deghosting algorithms for multi-exposure high dynamic range imaging. In *Spring Conference on Computer Graphics*, pages 21–28. ACM, 2013.

[44] K. K. Hadziabdic, J. H. Telalović, and R. Mantiuk. Expert evaluation of deghosting algorithms for multi-exposure high dynamic range imaging. In *HDRi2014 - Second International Conference and SME Workshop on HDR imaging*, Sarajevo, Bosnia and Herzegovina, April 2014.

[45] D. Hafner, O. Demetz, and J. Weickert. Simultaneous HDR and optic flow computation. In *Proceedings of 22nd International Conference on Pattern Recognition (ICPR 2014)*. IEEE, 2014.

[46] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, volume 15, page 50. Manchester, UK, 1988.

[47] Y. S. Heo, K. M. Lee, S. U. Lee, Y. Moon, and J. Cha. Ghost-free high dynamic range imaging. In *Computer Vision–ACCV*, pages 486–500. Springer, 2010.

[48] R. Herzog, M. Čadík, T. O. Aydın, K. I. Kim, K. Myszkowski, and H.-P. Seidel. NoRM: no-reference image quality metric for realistic image synthesis. *Computer Graphics Forum*, 31(2):545–554, 2012.

[49] I. Hossain and B. K. Gunturk. High dynamic range imaging of non-static scenes. In *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2011.

[50] J. Hu, O. Gallo, and K. Pulli. Exposure stacks of live scenes with hand-held cameras. In *European Conference on Computer Vision*, pages 499–512. Springer, 2012.

[51] J. Hu, O. Gallo, K. Pulli, and X. Sun. HDR deghosting: How to deal with saturation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.

[52] Hugin. Panorama photo stitcher. Accessed: September 2014.

[53] J. Im, S. Jang, S. Lee, and J. Paik. Geometrical transformation-based ghost artifacts removing for high dynamic range image. In *18th IEEE International Conference on Image Processing (ICIP)*, pages 357–360. IEEE, 2011.

[54] J. Im, S. Lee, and J. Paik. Improved elastic registration for removing ghost artifacts in high dynamic imaging. *IEEE Transactions on Consumer Electronics*, 57(2):932–935, 2011.

[55] L. Ingber et al. Adaptive simulated annealing (ASA): Lessons learned. *Control and cybernetics*, 25:33–54, 1996.

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

[57] K. Jacobs, C. Loscos, and G. Ward. Automatic high-dynamic range image generation for dynamic scenes. *IEEE Computer Graphics and Applications*, 28(2):84–93, 2008.

[58] T. Jinno and M. Okuda. Multiple exposure fusion for high dynamic range image acquisition. *IEEE Transactions on Image Processing*, 21(1):358–365, 2012.

[59] N. K. Kalantari, E. Shechtman, C. Barnes, S. Darabi, D. B. Goldman, and P. Sen. Patch-based high dynamic range video. *ACM Transactions Graphics*, 32(6), 2013.

[60] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High dynamic range video. *ACM Transactions on Graphics*, 22(3):319–325, 2003.

[61] W.-C. Kao, C.-C. Hsu, L.-Y. Chen, C.-C. Kao, and S.-H. Chen. Integrating image fusion and motion stabilization for capturing still images in high dynamic range scenes. *IEEE Transactions on Consumer Electronics*, 52(3):735–741, 2006.

[62] E. A. Khan, A. O. Akyüz, and E. Reinhard. Ghost removal in high dynamic range images. In *2006 IEEE International Conference on Image Processing*, pages 2005–2008. IEEE, 2006.

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

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

[65] C. L. Lawson and R. J. Hanson. *Solving least squares problems*, volume 161. SIAM, 1974.

[66] Le Kang, Peng Ye, Yi Li, and David Doermann. Convolutional Neural Networks for No-Reference Image Quality Assessment. In *IEEE Conf. On Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[67] C. Lee, Y. Li, and V. Monga. Ghost-free high dynamic range imaging via rank minimization. *IEEE Signal Processing Letters*, 21(9):1045–1049, Sept 2014.

[68] D.-K. Lee, R.-H. Park, and S. Chang. Improved histogram based ghost removal in exposure fusion for high dynamic range images. In *IEEE 15th International Symposium on Consumer Electronics (ISCE)*, pages 586–591. IEEE, 2011.

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

[70] Z. Li, S. Rahardja, Z. Zhu, S. Xie, and S. Wu. Movement detection for the synthesis of high dynamic range images. In *17th IEEE International Conference on Image Processing (ICIP)*, pages 3133–3136. IEEE, 2010.

[71] H.-Y. Lin and W.-Z. Chang. High dynamic range imaging for stereoscopic scene representation. In *16th IEEE International Conference on Image Processing (ICIP)*, pages 4305–4308, 2009.

[72] H. Liu and I. Heynderickx. Issues in the design of a no-reference metric for perceived blur. *Proceedings of SPIE*, 7867:78670C–78670C–8, 2011.

[73] Y. Liu, J. Wang, S. Cho, A. Finkelstein, and S. Rusinkiewicz. A no-reference metric for evaluating the quality of motion deblurring. *ACM Transactions Graphics*, 32(6):175:1–175:12, Nov. 2013.

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

[75] J. Lubin. A visual discrimination model for imaging system design and evaluation. In E. Peli, editor, *Vision Models for Target Detection and Recognition*, pages 245–283. World Scientific, 1995.

[76] B. D. Lucas, T. Kanade, et al. An iterative image registration technique with an application to stereo vision. In *IJCAI*, volume 81, pages 674–679, 1981.

[77] H. Malm, M. Oskarsson, E. Warrant, P. Clarberg, J. Hasselgren, and C. Lejd-fors. Adaptive enhancement and noise reduction in very low light-level video. In *IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE, 2007.

[78] H. Malm and E. Warrant. Motion dependent spatiotemporal smoothing for noise reduction in very dim light image sequences. In *18th International Conference on Pattern Recognition*, volume 3, pages 954–959. IEEE, 2006.

[79] S. Mangiat and J. Gibson. High dynamic range video with ghost removal. In *SPIE Optical Engineering+ Applications*. International Society for Optics and Photonics, 2010.

[80] S. Mangiat and J. Gibson. Spatially adaptive filtering for registration artifact removal in hdr video. In *18th IEEE International Conference on Image Processing*, pages 1317–1320. IEEE, 2011.

[81] S. Mann, C. Manders, and J. Fung. Painting with looks: Photographic images from video using quantimetric processing. In *Proceedings of the 10th ACM International Conference on Multimedia*, pages 117–126. ACM, 2002.

[82] S. Mann and R. W. Picard. On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. Technical Report 323, M.I.T. Media Lab Perceptual Computing Section, Boston, Massachusetts, 1994. also appears IS&T's 48th annual conference, Cambridge, Massachusetts, May 1995.

[83] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Transactions Graphics*, 30(4):40:1–40:14, July 2011.

[84] Y. Matsushita, E. Ofek, W. Ge, X. Tang, and H.-Y. Shum. Full-frame video stabilization with motion inpainting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(7):1150–1163, 2006.

[85] J. W. Mauchly. Significance test for sphericity of a normal n-variate distribution. *The Annals of Mathematical Statistics*, 11(2):204–209, 1940.

[86] N. Menzel and M. Guthe. Freehand HDR photography with motion compensation. In *VMV*, pages 127–134, 2007.

[87] T. Mertens, J. Kautz, and F. Van Reeth. Exposure fusion. In *15th Pacific Conference on Computer Graphics and Application*, pages 382–390. IEEE, 2007.

[88] T.-H. Min, R.-H. Park, and S. Chang. Histogram based ghost removal in high dynamic range images. In *IEEE International Conference on Multimedia and Expo*, pages 530–533. IEEE, 2009.

[89] T.-H. Min, R.-H. Park, and S. Chang. Noise reduction in high dynamic range images. *Signal, Image and Video Processing*, 5(3):315–328, 2011.

[90] T. Mitsunaga and S. K. Nayar. Radiometric self calibration. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1. IEEE, 1999.

[91] A. Mittal, A. K. Moorthy, and A. C. Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012.

[92] Y.-S. Moon, Y.-M. Tai, J. H. Cha, and S.-H. Lee. A simple ghost-free exposure fusion for embedded HDR imaging. In *IEEE International Conference on Consumer Electronics (ICCE)*, pages 9–10. IEEE, 2012.

[93] M. Narwaria, R. K. Mantiuk, M. P. Da Silva, and P. Le Callet. HDR-VDP-2.2: A calibrated method for objective quality prediction of high-dynamic range and standard images. *Journal of Electronic Imaging*, 24(1):010501, 2015.

[94] T. H. Oh, J.-Y. Lee, and I.-S. Kweon. High dynamic range imaging by a rank-1 constraint. In *ICIP*, pages 790–794, 2013.

[95] R. R. Orozco, I. Martín, C. Loscos, and P.-P. Vasquez. Full high-dynamic range images for dynamic scenes. In *SPIE Photonics Europe*. International Society for Optics and Photonics, 2012.

[96] A. Oğuz Akyüz and A. Genctav. A reality check for radiometric camera response recovery algorithms. *Computers & Graphics*, 37(7):935–943, 2013.

[97] S.-C. Park, H.-H. Oh, J.-H. Kwon, W. Choe, and S.-D. Lee. Motion artifact-free HDR imaging under dynamic environments. In *18th IEEE International Conference on Image Processing (ICIP)*, pages 353–356. IEEE, 2011.

[98] E. Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, pages 1065–1076, 1962.

[99] F. Pece and J. Kautz. Bitmap movement detection: HDR for dynamic scenes. In *Conference on Visual Media Production (CVMP)*, pages 1–8. IEEE, 2010.

[100] M. Pedone and J. Heikkilä. Constrain propagation for ghost removal in high dynamic range images. In *VISAPP*, pages 36–41, 2008.

[101] S. Periaswamy and H. Farid. Elastic registration in the presence of intensity variations. *IEEE Transactions on Medical Imaging*, 22(7):865–874, 2003.

[102] A. A. Rad, L. Meylan, P. Vandewalle, and S. Süsstrunk. Multidimensional image enhancement from a set of unregistered and differently exposed images. In *Electronic Imaging*. International Society for Optics and Photonics, 2007.

92

[103] S. Raman and S. Chaudhuri. Bottom-up segmentation for ghost-free reconstruction of a dynamic scene from multi-exposure images. In *Proceedings of the 7th Indian Conference on Computer Vision, Graphics and Image Processing*, pages 56–63. ACM, 2010.

[104] S. Raman and S. Chaudhuri. Reconstruction of high contrast images for dynamic scenes. *The Visual Computer*, 27(12):1099–1114, 2011.

[105] S. Raman, V. Kumar, and S. Chaudhuri. Blind de-ghosting for automatic multi-exposure compositing. In *SIGGRAPH ASIA Posters*, page 44. ACM, 2009.

[106] G. Ramanarayanan, J. Ferwerda, B. Walter, and K. Bala. Visual equivalence: Towards a new standard for image fidelity. *ACM Transactions on Graphics*, 26(3), July 2007.

[107] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda. Photographic tone reproduction for digital images. *ACM Transactions on Graphics*, 21(3):267–276, 2002.

[108] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec. *High Dynamic Range Imaging: Acquisition, Display and Image-Based Lighting*. Morgan Kaufmann, San Francisco, second edition edition, 2010.

[109] M. M. O. B. B. Richard and M. Y.-S. Chang. Fast digital image inpainting. In *Proceedings of the International Conference on Visualization, Imaging and Image Processing (VIIP)*, pages 106–107, 2001.

[110] M. A. Robertson, S. Borman, and R. L. Stevenson. Dynamic range improvement through multiple exposures. In *International Conference on Image Processing*, volume 3, pages 159–163. IEEE, 1999.

[111] M. Rosenblatt et al. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837, 1956.

[112] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir. A comparative study of image retargeting. *ACM Transactions on Graphics*, 29(6), 2010.

[113] P. Sand and S. Teller. Video matching. *ACM Transactions on Graphics*, 23(3):592–599, 2004.

[114] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs. High dynamic range display systems. *ACM Transactions on Graphics*, 23(3):760–768, 2004.

[115] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman. Robust patch-based HDR reconstruction of dynamic scenes. *ACM Transactions Graphics*, 31(6):203, 2012.

[116] R. H. L. Shapiro and R. Haralick. Computer and robot vision. *Reading: Addison-Wesley*, 1992.

[117] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Trans Image Processing*, 15(2):430–444, 2006.

[118] H. R. Sheikh, A. C. Bovik, and L. K. Cormack. No-reference quality assessment using natural scene statistics: JPEG2000. *IEEE Transactions on Image Processing*, 14(11):1918–1927, 2005.

[119] N. V. Shirahatti and K. Barnard. Evaluating image retrieval. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 955–961. IEEE, 2005.

[120] D. D. Sidibe, W. Puech, O. Strauss, et al. Ghost detection and removal in high dynamic range images. In *European Signal Processing Conference*, 2009.

[121] S. Silk and J. Lang. Fast high dynamic range image deghosting for arbitrary scene motion. In *Proceedings of Graphics Interface*, pages 85–92. Canadian Information Processing Society, 2012.

[122] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani. Summarizing visual data using bidirectional similarity. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[123] E. P. Simoncelli. Statistical modeling of photographic images. In A. Bovik, editor, *Handbook of Image and Video Processing*, chapter 4.7, pages 431–441. Academic Press, May 2005. 2nd edition.

[124] K. Smith, G. Krawczyk, K. Myszkowski, and H.-P. Seidel. Beyond tone mapping: Enhanced depiction of tone mapped hdr images. *Comput. Graph. Forum*, 25(3):427–438, 2006.

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

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

[127] A. Srikantha, D. Sidibé, and F. Mériaudeau. An SVD-based approach for ghost detection and removal in high dynamic range images. In *21st International Conference on Pattern Recognition (ICPR)*, pages 380–383. IEEE, 2012.

[128] T. Starks and H. David. Significance tests for paired-comparison experiments. *Biometrika*, pages 95–108, 1961.

[129] H.-S. Sung, R.-H. Park, D.-K. Lee, and S. Chang. Feature based ghost removal in high dynamic range imaging. *International Journal of Computer Graphics & Animation*, 3(4), 2013.

[130] H. Tang, N. Joshi, and A. Kapoor. Learning a blind measure of perceptual image quality. In *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[131] P. Teo and D. Heeger. Perceptual image distortion. In *IEEE International Conference on Image Processing*. IEEE, 1994.

[132] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *6th International Confereance on Computer Vision*, pages 839–846. IEEE, 1998.

[133] A. Tomaszewska and R. Mantiuk. Image registration for multi-exposure high dynamic range image acquisition. In *15th International Conference Central Europe on Computer Graphics, Visualization and Computer Vision*, 2007.

[134] F. Tombari, L. Di Stefano, and S. Mattoccia. A robust measure for visual correspondence. In *14th International Conference on Image Analysis and Processing*, pages 376–381. IEEE, 2007.

[135] O. T. Tursun, A. O. Akyuz, A. Erdem, and E. Erdem. Evaluating deghosting algorithms for HDR images. In *22nd Signal Processing and Communications Applications Conference (SIU)*, pages 1275–1278. IEEE, 2014.

[136] O. T. Tursun, A. O. Akyüz, A. Erdem, and E. Erdem. The state of the art in hdr deghosting: A survey and evaluation. *Computer Graphics Forum*, 34(2):683–707, 2015.

[137] C. Wang and C. Tu. An exposure fusion approach without ghost for dynamic scenes. In *6th International Congress on Image and Signal Processing*, volume 2, pages 904–909. IEEE, 2013.

[138] Z. Wang and A. C. Bovik. *Modern Image Quality Assessment*. Morgan and Claypool Publishers, 2006.

[139] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[140] G. Ward. Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. *Journal of graphics tools*, 8(2):17–30, 2003.

[141] H. R. Wu and K. R. Rao. *Digital Video Image Quality and Perceptual Coding (Signal Processing and Communications)*. CRC Press, Inc., Boca Raton, FL, USA, 2005.

[142] S. Wu, S. Xie, S. Rahardja, and Z. Li. A robust and fast anti-ghosting algorithm for high dynamic range imaging. In *17th IEEE International Conference on Image Processing (ICIP)*, pages 397–400. IEEE, 2010.

[143] G.-Z. Yang, P. Burger, D. N. Firmin, and S. Underwood. Structure adaptive anisotropic image filtering. *Image and Vision Computing*, 14(2):135–145, 1996.

[144] S. Yao. Robust image registration for multiple exposure high dynamic range image synthesis. In *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2011.

[145] P. Ye, J. Kumar, L. Kang, and D. Doermann. Unsupervised feature learning framework for no-reference image quality assessment. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[146] Y. Yoo, K. Lee, W. Choe, S. Park, S.-D. Lee, and C.-Y. Kim. A digital ISO expansion technique for digital cameras. In *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2010.

[147] C. Zach, T. Pock, and H. Bischof. A duality based approach for realtime TV-L 1 optical flow. In *Pattern Recognition*, pages 214–223. Springer, 2007.

[148] L. Zhang, D. Zhang, X. Mou, and D. Zhang. FSIM:a feature similarity index for image quality assessment. *IEEE Trans Image Processing*, 20(8):2378–2386, 2011.

[149] W. Zhang and W.-K. Cham. Gradient-directed composition of multi-exposure images. In *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 530 –536, june 2010.

[150] W. Zhang and W.-K. Cham. Gradient-directed multiexposure composition. *IEEE Transactions on Image Processing*, 21(4):2318–2323, 2012.

[151] W. Zhang and W.-K. Cham. Reference-guided exposure fusion in dynamic scenes. *Journal of Visual Communication and Image Representation*, 23(3):467–475, 2012.

[152] J. Zheng, Z. Li, Z. Zhu, and S. Rahardja. A hybrid patching scheme for high dynamic range imaging. In *Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, 2011.

[153] J. Zheng, Z. Li, Z. Zhu, S. Wu, and S. Rahardja. Patching of moving objects for ghosting-free hdr synthesis. In *ACM SIGGRAPH Posters*, page 62. ACM, 2012.

[154] J. Zheng, Z. Li, Z. Zhu, S. Wu, and S. Rahardja. Hybrid patching for a sequence of differently exposed images with moving objects. *IEEE Transactions on Image Processing*, 22(12):5190–5201, 2013.

[155] H. Zimmer, A. Bruhn, and J. Weickert. Freehand HDR imaging of moving scenes with simultaneous resolution enhancement. *Computer Graphics Forum*, 30(2):405–414, 2011.

APPENDIX A

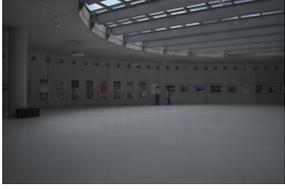# DATASET FROM THE FIRST SUBJECTIVE EXPERIMENT

Scene	Input EV -1	Input EV 0	Input EV +1
Cafe			
Candles			
FastCars			
Flag			
Gallery1			

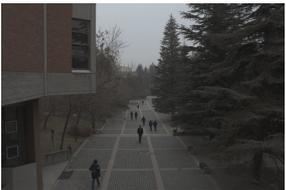Figure A.1: Representative images for the scenes from the first subjective experiment

Scene	Input EV -1	Input EV 0	Input EV +1
Gallery2			
LibrarySide			
Shop1			
Shop2			
WalkingPeople			
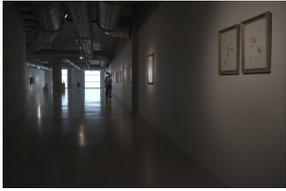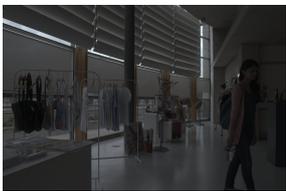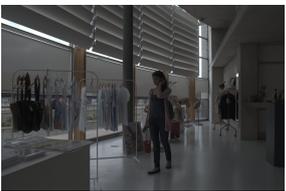
Figure A.1 (cont.): Representative images for the scenes from the first subjective experiment

APPENDIX B

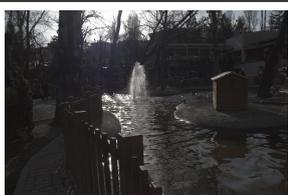# OUTPUTS OF HDR DEGHOSTING ALGORITHMS FOR THE FIRST DATASET

Scene	Grosch [38]	Khan et al. [62]	Sen et al. [115]
Cafe			
Candles			
FastCars			
Flag			

Figure B.1: Outputs of Grosch [38], Khan et al. [62] and Sen et al. [115] for the first dataset

Scene	Grosch [38]	Khan et al. [62]	Sen et al. [115]
Gallery1			
Gallery2			
LibrarySide			
Shop1			
Shop2			
WalkingPeople			

Figure B.1 (cont.): Outputs of Grosch [38], Khan et al. [62] and Sen et al. [115] for the first dataset

Scene	Silk&Lang-FM [121]	Hu et al. [51]	SimpleDG
Cafe			
Candles			
FastCars			
Flag			
Gallery1			
Gallery2			
LibrarySide			
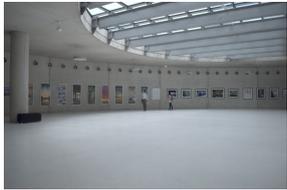
Figure B.2: Outputs of Silk&Lang-FM [121], Hu et al. [51], SimpleDG for the first dataset

Scene	Silk&Lang-FM [121]	Hu et al. [51]	SimpleDG
Shop1			
Shop2			
WalkingPeople			

Figure B.2 (cont.): Outputs of Silk&Lang-FM [121], Hu et al. [51], SimpleDG for the first dataset

APPENDIX C

DATASET FROM THE SECOND SUBJECTIVE EXPERIMENT

Scene	Input EV -1	Input EV 0	Input EV +1
StreetNight			
Museum1			
Museum2			
Museum3			
StreetDay			
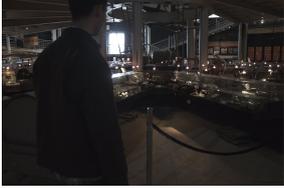
Figure C.1: Representative images for the scenes from the first subjective experiment

Scene	Input EV -1	Input EV 0	Input EV +1
Plants			
Exit			
Pendulum			
Cars			
Pedestrians			
Fountain			
Yard			
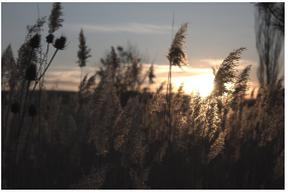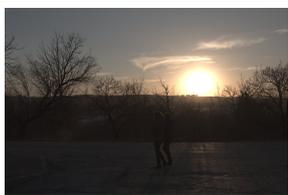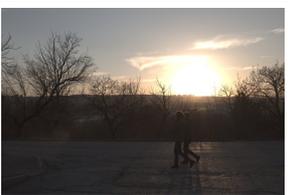
Figure C.1 (cont.): Representative images for the scenes from the first subjective experiment

Scene	Input EV -1	Input EV 0	Input EV +1
Turtles			
Building			
Flames			
Toytrain			
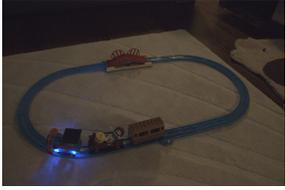
Figure C.1 (cont.): Representative images for the scenes from the first subjective experiment

APPENDIX D

OUTPUTS OF HDR DEGHOSTING ALGORITHMS FOR THE SECOND DATASET

Scene	Grosch [38]	Khan et al. [62]	Sen et al. [115]
StreetNight			
Museum1			
Museum2			
Museum3			

Figure D.1: Outputs of Grosch [38], Khan et al. [62] and Sen et al. [115] for the second dataset

Scene	Grosch [38]	Khan et al. [62]	Sen et al. [115]
StreetDay			
Plants			
Exit			
Pendulum			
Cars			
Pedestrians			
Fountain			
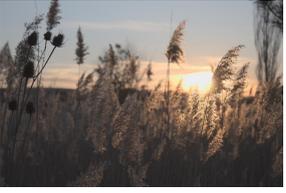
Figure D.1 (cont.): Outputs of Grosch [38], Khan et al. [62] and Sen et al. [115] for the second dataset

Scene	Grosch [38]	Khan et al. [62]	Sen et al. [115]
Yard			
Turtles			
Building			
Flames			
Toytrain			

Figure D.1 (cont.): Outputs of Grosch [38], Khan et al. [62] and Sen et al. [115] for the second dataset

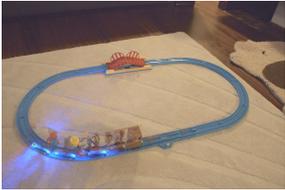Figure D.2: Outputs of Silk&Lang-FM [121], Silk&Lang-PWD [121], Srikantha et al. [127] for the second dataset

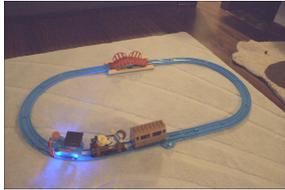| Scene	Silk&Lang-FM [121]	Silk&Lang-PWD [121]	Srikantha et al. [127]
Pendulum			
Cars			
Pedestrians			
Fountain			
Yard			
Turtles			
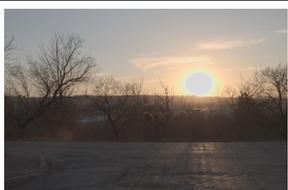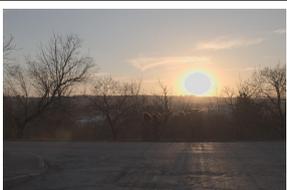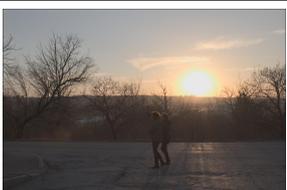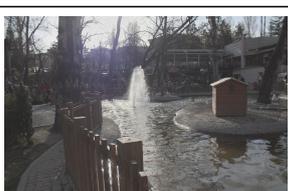
Figure D.2 (cont.): Outputs of Silk&Lang-FM [121], Silk&Lang-PWD [121], Srikantha et al. [127] for the second dataset

Scene	Silk&Lang-FM [121]	Silk&Lang-PWD [121]	Srikantha et al. [127]
Building			
Flames			
Toytrain			

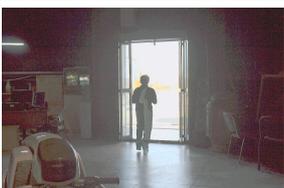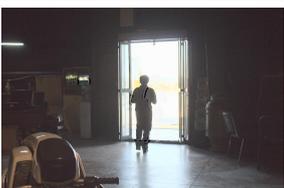Figure D.2 (cont.): Outputs of Silk&Lang-FM [121], Silk&Lang-PWD [121], Srikantha et al. [127] for the second dataset

Scene	Hu et al. [51]	Lee et al. [67]	SimpleDG
StreetNight			
Museum1			
Museum2			
Museum3			
StreetDay			
Plants			
Exit			

Figure D.3: Outputs of Hu et al. [51], Lee et al. [67], SimpleDG for the second dataset

Scene	Hu et al. [51]	Lee et al. [67]	SimpleDG
Pendulum			
Cars			
Pedestrians			
Fountain			
Yard			
Turtles			
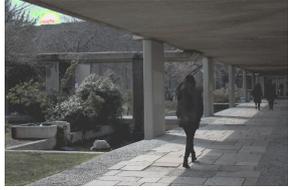
Figure D.3 (cont.): Outputs of Hu et al. [51], Lee et al. [67], SimpleDG for the second dataset

Scene	Hu et al. [51]	Lee et al. [67]	SimpleDG
Building			
Flames			
Toytrain			
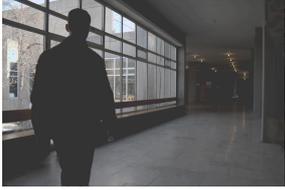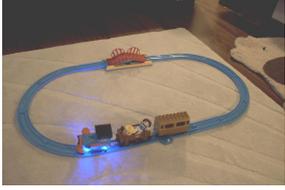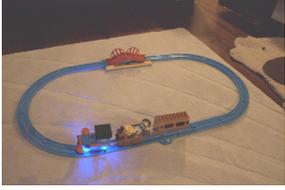
Figure D.3 (cont.): Outputs of Hu et al. [51], Lee et al. [67], SimpleDG for the second dataset

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Tursun, Okan Tarhan
Nationality: Turkish (TC)
Date and Place of Birth: 1986, Bandırma
Marital Status: Single

EDUCATION

Degree	Institution	Year of Graduation
Double Major	Dept. of Computer Engineering, METU	2010
B.S.	Dept. of Industrial Engineering, METU	2009

PROFESSIONAL EXPERIENCE

Year	Place	Enrollment
Sept. 2010-Feb. 2016	Dept. of Computer Engineering	Teaching Assistant

PUBLICATIONS

[1] Okan Tarhan Tursun, Ahmet Oğuz Akyüz, Aykut Erdem, and Erkut Erdem. An Objective Deghosting Quality Metric for HDR Images. To appear in Computer Graphics Forum (Proc. of EG2016).

[2] Okan Tarhan Tursun, Ahmet Oğuz Akyüz, Aykut Erdem, and Erkut Erdem. The state of the art in HDR deghosting: A survey and evaluation. Computer Graphics Forum, 34(2):683–707, 2015.

[3] Ahmet Oğuz Akyüz, Okan Tarhan Tursun, Jasminka Hasić-Telalović, and Kanita Karađuzović-Hadžiabdić. HDR Video Capture and Deghosting (Book chapter - in preparation). Elsevier, 2015.

[4] Okan Tarhan Tursun, Ahmet Oğuz Akyüz, Aykut Erdem, and Erkut Erdem. Evaluating deghosting algorithms for HDR images. In Signal Processing and Communications Applications Conference (SIU), 2014 22nd, page 1275–1278. IEEE, 2014.