On the Generation of High Dynamic Range Images: Theory and Practice from a Statistical Perspective

Cecilia Aguerrebere

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2014-ENST-0026

UNIVERSIDAD DE LA REPUBLICA URUGUAY

FACULTAD DE INGENIERIA

EDITE - ED 130

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech et Universidad de la República, Uruguay

Spécialité « SIGNAL et IMAGES »

Cecilia AGUERREBERE OTEGUI

Soutenue le 28 Mai 2014

Sur la génération d'images à grande gamme dynamique

Théorie et pratique : une perspective statistique

Jury

Lionel MOISAN	Université Paris Descartes	Rapporteur
Gabriel PEYRÉ	Université Paris-Dauphine	Rapporteur
Antoni BUADES	Ecole Normale Supérieure de Cachan	Examinateur
Frédéric GUICHARD	DxO Labs	Examinateur
Joseph SALMON	Télécom ParisTech	Examinateur
Julie DELON	Université Paris Descartes	Directeur
Yann GOUSSEAU	Télécom ParisTech	Directeur
Pablo MUSÉ	Universidad de la República, Uruguay	Directeur
Henri MAÎTRE	Télécom ParisTech	Invité

TELECOM ParisTech
école de l'Institut Télécom - membre de ParisTech

T H È S E

A mis viejos Pilar y Sergio.
A Pau, Leti y Fede.

Abstract

This dissertation studies the problem of high dynamic range (HDR) image generation from a statistical perspective. A thorough analysis of the camera acquisition process leads to a simplified yet realistic statistical model describing raw pixel values. The analysis and methods then proposed are based on this model.

First, the theoretical performance bound of the problem is computed for the static case, where the acquisition conditions are controlled. Furthermore, a new method is proposed that, unlike previous methods, improves the reconstructed HDR image by taking into account the information carried by saturated samples.

From a more practical perspective, two methods are proposed to generate HDR images in the more realistic and complex case where both objects and camera may exhibit motion. The first one is a multi-image, patch-based method, that simultaneously estimates and denoises the HDR image. The other is a single image approach that makes use of a general restoration method to generate the HDR image. This general restoration method, applicable to a wide range of problems, constitutes the last contribution of this dissertation.

Résumé

Cette thèse porte sur le problème de la génération d'images à grande gamme dynamique (HDR pour l'anglais High Dynamic Range). Une analyse approfondie du processus d'acquisition de la caméra conduit tout d'abord à un modèle statistique simplifié mais réaliste décrivant les valeurs brutes des pixels. Les analyses et méthodes proposées par la suite sont fondées sur ce modèle.

Nous posons le problème de l'estimation de l'irradiance comme un problème d'estimation statistique et en calculons la borne de performance. Les performances des estimateurs d'irradiance classiques sont comparées à cette borne. Les résultats obtenus justifient l'introduction d'un nouvel estimateur qui, au contraire des méthodes de la littérature, prend en compte les échantillons saturés.

D'un point de vue plus pratique, deux méthodes sont proposées pour générer des images HDR dans le cas plus réaliste et complexe de scènes dynamiques. Nous proposons tout d'abord une méthode multi-image qui utilise des voisinages (*patches*) pour estimer et débruiter l'image HDR de façon simultanée. Nous proposons également une approche qui repose sur l'acquisition d'une seule image. Cette approche repose sur une méthode générique, par *patches*, de résolution des problèmes inverses pour génerer l'image HDR. Cette méthode de restauration, d'un point de vue plus général et pour une large gamme d'applications, constitue la dernière contribution de cette thèse.

Resumen

Esta tesis estudia el problema de la generación de imágenes de amplio rango dinámico (HDR por sus sigla en inglés: high dynamic range) desde una perspectiva de análisis estadístico. El estudio en profundidad del proceso de adquisición de una cámara digital nos permite describir los valores de los píxeles a través de un modelo estadístico simple y realista. Este modelo sirve de base tanto para el análisis teórico como para los métodos prácticos presentados en esta tesis.

La cota teórica de máximo desempeño para este problema es calculada en el caso estático, donde las condiciones de adquisición son controladas. Este análisis motiva la creación de un nuevo método que, a diferencia de los métodos precedentes, mejora la calidad de la imagen reconstruida haciendo uso de la información provista por las muestras saturadas.

Finalmente, desde un punto de vista más práctico, dos métodos son propuestos para generar imágenes HDR en el caso más realista y por tanto más complejo dado por las escenas dinámicas. El primero es un método multi-imagen, que utiliza patches para realizar de forma simultánea la estimación de la imagen HDR y la remoción de ruido en la misma. El segundo utiliza una única imagen y hace uso de un método general de reconstrucción para recomponer la información de amplio rango dinámico. Este método general de restauración, aplicable a un rango más amplio de problemas, constituye la última contribución de esta tesis.

Contents

1 Introduction **13**

1.1 Camera model 18
1.2 HDR imaging for static scenes and static camera 19
1.3 HDR imaging for dynamic scenes and hand-held camera 21
1.4 A general image restoration method 26

2 Camera acquisition model **29**

2.1 Acquisition of digital images 29
2.2 Camera calibration procedure 36

3 HDR imaging for static scenes and static camera **39**

3.1 Previous work 41
3.2 Performance Bounds 43
3.3 Including saturation information 58
3.4 Model parameters uncertainties and performance bounds 67
3.5 Conclusions 71

4 Simultaneous HDR and super-resolution **73**

4.1 HDR-SR acquisition and reconstruction 75
4.2 Strategy for the choice of acquisition parameters 76
4.3 Reconstruction error bound for HDR-SR 78
4.4 Exposure times selection for HDR reconstruction 82
4.5 Acquisition strategy 84
4.6 Conclusions 89

5 HDR imaging for dynamic scenes and hand-held camera **93**

5.1 Previous work 93
5.2 HDR imaging for dynamic scenes and hand-held camera 95
5.3 Results 101
5.4 Conclusions 108

6 A general restoration method with an application to single-image HDR 111
6.1 Piecewise linear estimators 113
6.2 Non local Bayes 116
6.3 Proposed approach: Hyperprior Non Local Bayes 118
6.4 Application: single image HDR 135
6.5 Conclusions 160

7 Conclusions and Perspectives 165

Appendices 169

A Tone mapping operators 171

B MAP estimation of parameters 173

Bibliography 177

1 Introduction

Imagine you are on holidays in some lost place. The sun is shining, the weather is sweet. Suddenly, you see the most beautiful scene you want to keep forever engraved in your memory. In the distance, the splendorous sun starts its countdown for sunset, still powerful enough to light the calm sea. Close, vegetation stands, setting up a leafy frame, ultimately reaching the mountain. The composition is just magnificent. You take your camera and shoot your best shot. And... no... it is just not possible. The camera and you are definitely not looking at the same scene. Otherwise, how could it give such a mediocre and inexpressive result? Well, that's when you need high dynamic range imaging.

High dynamic range (HDR) imaging is a field of image processing that aims at reproducing an extended dynamic range of luminosity compared with that which can be captured using a standard digital camera. The range of luminosity which a standard digital camera can capture is often not enough to produce a faithful representation of real scenes. High dynamic range imaging techniques aim at narrowing the gap between what our eyes can see and what can be registered in an image.

In the last few years, as a result of the widespread use of digital cameras, in particular those built-into smart-phones, HDR images have become widely known and available for the general public. The creation of HDR images is an automatic function in most modern digital single-lens reflex (SLR) cameras and can be enabled in most smart-phones simply by installing a specific application. High dynamic range imaging allows to go one step further in the search for a faithful representation of real world scenes. With very simple means, as a digital camera is nowadays, it allows the general public to create outstanding photographs that would be unattainable some years ago.

But why standard digital cameras cannot capture the scenes as we see them? And how HDR imaging techniques aim at overcoming this obstacle?

The photon bucket limitation The human visual system can capture scenes of very high dynamic range, correctly perceiving details in both dark and bright regions. It can adapt simultaneously to a range of intensity levels in the order of 17 stops[1] [Reinhard et al. 2010]. This is not the case for current standard digital cameras, which can reach up to 14.4 stops

[1]The dynamic range measured in stops is equal to the base-2 logarithm of the ratio between the brightest and the darkest considered values.

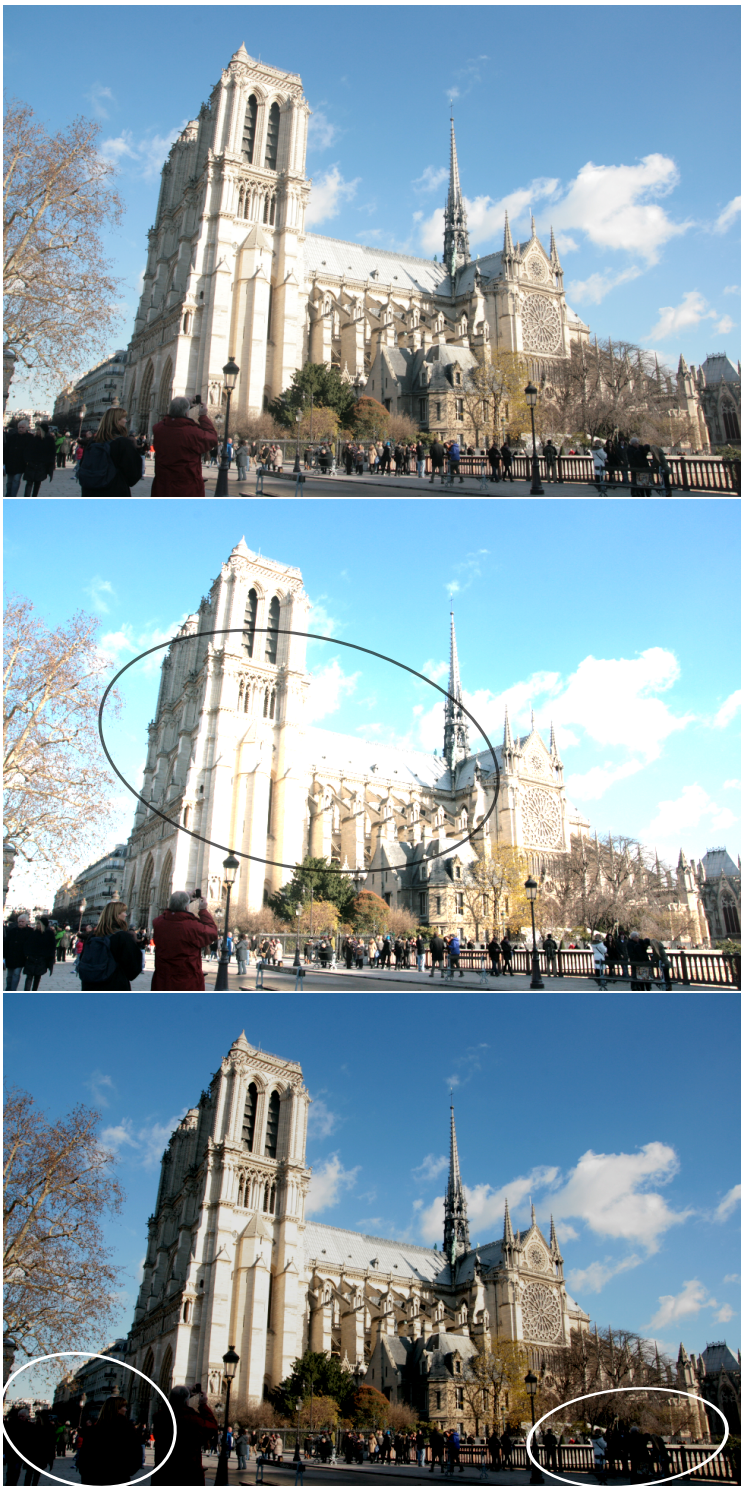


Figure 1.1: **Top:** High dynamic range (HDR) image of a HDR scene. **Center:** Photograph of a HDR scene captured with a standard digital camera using a long exposure time. Details are lost in the bright regions due to saturation. **Bottom:** Photograph of a HDR scene captured with a standard digital camera using a short exposure time. Details are lost in the bright regions due to under-exposure. Details in both bright and dark regions cannot be kept with a single shot of a HDR scene using a standard digital camera.

Condition	Illumination (candela/m^2)
starlight	10^{-3}
moonlight	10^{-1}
indoor lighting	10^{2}
sunlight	10^{5}
maximum intensity of common CRT monitors	10^{2}

Table 1.1: Typical ambient luminance levels for natural scenes (from [Reinhard et al. 2010]).

for modern SLR cameras (e.g., Nikon D800 [DxO Labs]). Each image sensor cell, or pixel, acts as a photon to electron converter, transforming all the incoming photons into electrons. We can think of it as a bucket, called potential well, in which the electrons are accumulated during a given period of time. That time, called shutter speed or exposure time, is the length of time during which the camera's shutter remains open. These electrons are then read out as a potential difference, and the corresponding voltage value is digitized to give the pixel value we obtain at the output. Neglecting noise sources, this is exactly what raw pixel values are: a value proportional to the product flux of photons per unit time (or irradiance) times the exposure time.

As with any sort of bucket, the potential well has a limited capacity. Hence, for the bright regions of a scene, where the irradiance is very high, if a long exposure time is used, the irradiance multiplied by the exposure time may become way too large to be stored in the potential well. In those cases, pixels saturate resulting in information loss under the form of censored data.

This problem can be easily solved by sufficiently reducing the exposure time. However, with this solution, a new problem arises for the dark regions of the scene. For dark regions, the irradiance may be so low that very few photons reach the pixel per unit time. Therefore, when using a short exposure, the number of photons, and hence of electrons, accumulated in the potential well is very small and will be masked by the various sensor noise sources. Table 1.1 gives typical ambient luminance levels for natural scenes [Reinhard et al. 2010].

Take for example the HDR scene at the top of Figure 1.1. A single shot of this scene captured with a long exposure gives the central image in Figure 1.1. The details in the bright region of the Notre Dame cathedral are completely lost due to saturation. On the other hand, if the exposure time is reduced in order to avoid saturation, the dark regions become under-exposed and details are lost, as shown in the bottom image in Figure 1.1. Therefore, the result of a single shot picture of an HDR scene, taken with a regular digital camera, contains pixels which are either saturated or under-exposed, and cannot represent the whole dynamic range of the scene.

The beginning of HDR As early as in 1856, photographers were dealing with this problem and started developing the first techniques that laid out the basis of what we know today as HDR imaging. The French photographer Gustave Le Gray, one of the pioneers in this subject, was one of the first to capture an HDR scene, composed by sea and sky, by combining two differently exposed negatives (Figure 1.2). This technique was known as

Figure 1.2: *Brig upon the water* by Gustave Le Gray (1856). One of the first known examples of HDR images, created by the combination of two differently exposed negatives using the technique known as *ciel rapporté*.

ciel rapporté, since one negative was used for the landscape and another for the sky. His technique, developed for analog photography, inspired the basic idea which is used today to create HDR digital images, that is, to combine images acquired with different exposures. To the best of our knowledge, this idea was first introduced in digital photography by Mann and Picard in 1995.

HDR imaging: the numerical age Nowadays, the most common way to create HDR images is to combine multiple low dynamic range (LDR) photographs, acquired with different exposure times $\tau_1, \tau_2, \ldots, \tau_T$. Indeed, as previously mentioned, for a given irradiance C and exposure time τ_i, the corresponding raw pixel value is a function of the received luminous energy $\tau_i C$. Hence, using different exposure times controls the amount of received luminous energy, avoiding saturation for at least some of the exposures.

This way of sampling the camera response function at different operating points make it possible to capture details in both dark and bright regions. Figure 1.3 illustrates this acquisition process, known as exposure bracketing. Of course, the more photographs spanning the whole range of exposure times, the better the result. Ground-truth HDR images are produced in such conditions, in a very controlled environment. However, for practical limitations, exposure bracketing in real situations can rarely exceed a few snapshots.

The problem faced by HDR image generation techniques is how to combine the differently exposed images to create a single image, representing the irradiance map, that correctly describes the whole dynamic range of the original scene. The characteristics of the scene and the conditions of acquisition determine the level of complexity of the problem to be solved. For instance, the scene may be completely static or present moving objects, and the camera may be fixed or hand-held.

The problem of imaging dynamic scenes using a hand-held camera is much more com-

plicated than the one of static scenes using a fixed camera. As for all multi-imaging techniques, motion is a hard obstacle to overcome. In most cases, images captured with a hand-held camera need to be realigned and moving objects need to be identified in order to compensate for motion. Otherwise, severe ghosting artifacts appear which are particularly annoying in a final result and unacceptable for most applications.

An alternative to multi-image HDR is to modify the image sensor so as to increase its dynamic range, which is typically defined as the ratio between the maximum achievable signal (the potential well full capacity) divided by the smallest detectable signal, which is determined by the sensor noise. Image sensor technology is constantly evolving and several efforts have been made so as to enlarge their dynamic range (see for example [Reinhard et al. 2010] for a description of recent wide dynamic range sensors technology). One possibility is noise reduction, so as to increase the potential well effective capacity. Another option is to improve the acquisition strategy, performing for instance a multiple-capture, where the sensor is filled and read out several times per acquisition period, or the so called time-based image sensors, which encode luminance information by keeping track of the time it takes a pixel to saturate instead of the output voltage [Guo et al. 2007; Kavusi and El Gamal 2004]. Nayar and Mitsunaga [2000] proposed a new acquisition method that consists in varying the pixel exposures by placing an optical mask with spatially varying transmittance on top of the conventional sensor. This technique, known as spatially varying pixel exposures (SVE), has been developed further in recent years and extended to the acquisition of multi-spectral images [Yasuma et al. 2010]. Nevertheless, these technologies are still almost exclusively for professional use for specific applications and they are not yet available to the general public.

Displaying HDR images High dynamic range imaging techniques are used to create HDR images. Another problem arises at the moment of displaying these images in a conventional display. Most existing displays provide a very limited contrast range if compared to those attainable by HDR imaging techniques. Therefore, a technique is needed to compress the dynamic range of the HDR image in order to correctly display it on an standard display. Such an operation is usually called tone mapping or HDR image rendering. A summary of the existing tone mapping operators is included in Appendix A. Even though HDR displays with extended contrast ranges exist, their use is not yet generalized and a tone mapping operator is usually crucial for displaying HDR images. Figure 1.4 shows a diagram of the pipeline of the HDR imaging process, from acquisition to display.

Outline This thesis focuses on the generation of HDR images. In the following chapters we will study the HDR imaging problem from different perspectives. First, we start with the simplest case, that is, the static case where both the camera and the scene are fixed. In that context, we analyze the creation of HDR images and also the combined problem of HDR imaging and super-resolution. Then, we continue with a more realistic case, that is the case of dynamic scenes, where several moving objects appear, and images are captured using a hand-held camera.

Finally, we study a more general reconstruction problem which is subsequently applied to an HDR acquisition strategy that makes use of a single image captured with spatially

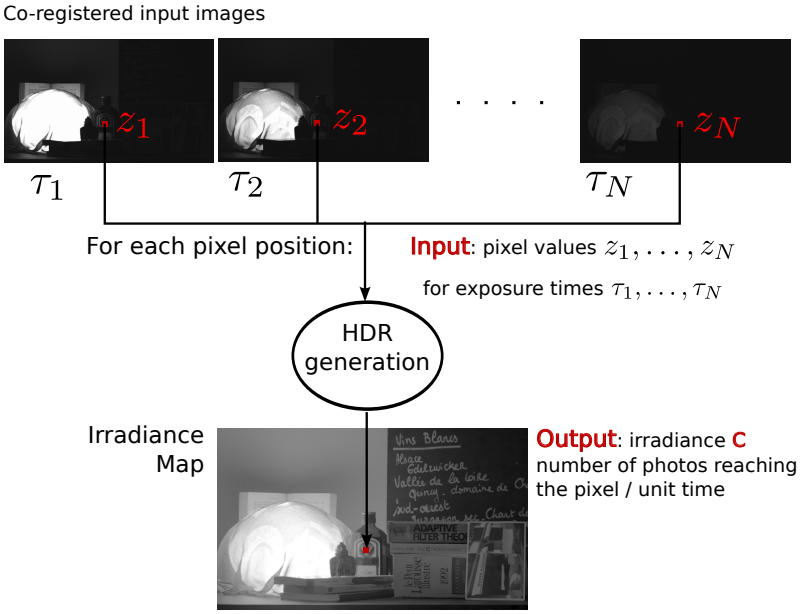


Figure 1.3: Diagram of the exposure bracketing process. Several images are captured using different exposure times. Bright regions are correctly captured with short exposures and dark regions with long exposures. These images are then combined into a single HDR image or irradiance map, where each pixel value represents the number of photons reaching that pixel per unit time.

varying pixel exposures.

The starting point of this work is the study of the digital camera acquisition process. All along this thesis, we will work with raw image data because a sound statistical noise model can be developed for it. Having a thorough understanding of the acquisition process and an accurate noise model is imperative to develop well established methods to tackle the aforementioned problems.

1.1 Camera model

After a thorough review of the acquisition process of digital cameras, from the photon capture in the sensor to the output of the digital pixel value, we propose a simplified yet realistic model describing the raw pixel values. This model takes into account the main noise sources: the Poisson photon shot noise due to the discrete nature of light, which can be approximated by a Gaussian distribution with equal mean and variance; the thermally generated readout noise, which can be modeled by a Gaussian distribution; the dark current shot noise; the spatial non-uniformities (photo-response non-uniformity (PRNU) and dark-current non-uniformity); and quantization noise. After considering the relative importance of the different noise sources we propose the following simplified model for non saturated nor under-exposed raw pixel values:

$$\mathrm{Z} \sim \mathcal{N}(ga\tau C + \mu_R, g^2 a\tau C + \sigma_R^2), \tag{1.1}$$

where g is the camera gain, a is the PRNU factor, τ is the exposure time, C is the irradiance reaching the pixel, μ_R and σ_R^2 are the readout noise mean and variance.

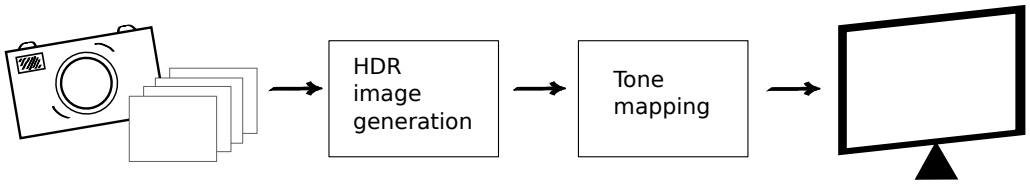


Figure 1.4: Pipeline of the HDR imaging process, from acquisition to display.

We will refer to this model several times throughout this work and will make use of it in order to develop rigorous solutions to the addressed problems. Since we work with raw data, we assume hereafter a linear camera response function. The camera parameters are assumed to be known, obtained by a calibration procedure. The only unknown in (1.1) is the irradiance C that we seek to estimate in order to compute the HDR image or irradiance map.

1.2 HDR imaging for static scenes and static camera

Mann and Picard [1995] considered the case where both the camera and the scene are fixed so that the images are perfectly aligned. They proposed to create an HDR image by computing the irradiance at each pixel p as a weighted average of the input samples $\mathrm{z}_1^p, \ldots, \mathrm{z}_T^p$ for that pixel,

$$\hat{C}_p = \frac{\sum_{i=1}^{T} w_i^p \frac{f^{-1}(z_i^p)}{\tau_i}}{\sum_{i=1}^{T} w_i^p}, \tag{1.2}$$

were w_i^p is the weight assigned to the exposure i for pixel p. They assign small weights to extreme pixel values – very low or close to saturation – based on the idea that the camera response function calibration is less accurate at these values. Saturated samples are discarded. Several methods followed which proposed different weighting schemes [Debevec and Malik 1997; Mitsunaga and Nayar 1999; Tsin et al. 2001; Reinhard et al. 2005; Robertson et al. 2003; Kirk and Andersen 2006; Granados et al. 2010]. A detailed description and evaluation of several of these methods is presented in Chapter 3.

Performance bounds In this work, we aim at answering the following questions: what is the performance bound of the HDR imaging problem in the static case? how far do existing methods perform from this bound?

Assuming that we have T independent samples per pixel, following Model (1.1), we compute the Cramér-Rao lower bound (CRLB) for unbiased estimators. This gives us a lower bound on the variance of any unbiased estimator of the irradiance computed from those samples. Applying the Cramér-Rao theorem (see, for instance [Kay 1993]), we show that this bound cannot be attained and therefore no efficient unbiased estimator exists for the irradiance under the considered hypotheses.

An experimental analysis is then conducted, where the performance of various existing irradiance estimators is compared to the CRLB. We find that the approximation of the maximum likelihood estimator (MLE) proposed by Granados et al. [2010], shown to be

the state-of-the-art [Granados et al. 2010], not only outperforms the other tested estimators but also has a nearly optimal behavior.

The MLE is known to be asymptotically efficient. However, the estimation is here performed with very few samples, thus the asymptotic assumption cannot be held. Some theoretical hints are outlined, explaining this a priori unexpected nearly optimal behavior for the MLE. Hence, we conclude that under the considered hypotheses, there is not much room for improvement for the pixel-wise irradiance estimation.

Including saturation information The comparison of the CRLB computed from all input samples with that computed when considering only non saturated samples, reveals the highly negative impact of saturation in the irradiance estimation performance. To the best of our knowledge, all existing irradiance estimators discard saturated samples. However, the fact that a pixel saturates for a given exposure time carries some amount of information with respect to the underlying irradiance value. Hence, part of the work presented in Chapter 3 is devoted to the study of the impact that saturation has in the irradiance estimation and whether it is possible to retrieve some information from saturated samples instead of just discarding them.

From this analysis, we propose a modified likelihood function adding a term that accounts for the probability of saturation of a given irradiance for a given exposure time. This modified likelihood is inspired by the work by Dempster et al. [1977] for parameter estimation in the case of censored data, which leads to the proposal of an expectation maximization (EM) procedure for the likelihood maximization.

We compute the CRLB for this modified likelihood function and show that the estimation performance can be considerably improved with the inclusion of saturation information for those irradiance values which are close to saturation.

Parameters uncertainty All existing irradiance estimation methods make use of some kind of pre-processing step where the camera response function is estimated and its inverse is computed. In this work, for instance, we assume a linear camera response function since we work with raw data, and we follow the procedure described in [Granados et al. 2010] to compute the camera parameters.

An important question, that does not seem to have drawn much attention, is the impact that the uncertainty of the camera response function estimation has in the performance of the estimators. Through an experimental analysis we show that an accurate camera calibration is crucial for the MLE to achieve its nearly optimal performance.

Furthermore, at a more fundamental level, we show that an accurate estimation of the camera parameters is fundamental in order to obtain a reliable ground-truth. In order to evaluate an HDR imaging technique, if the ground-truth is to be built using a standard digital camera (not a camera able to directly capture HDR information), an HDR method must be used to compute it from the captured images. If the camera calibration is not accurate enough, the ground-truth obtained in this manner will be biased and the estimation results might be judged better than what they actually are. This analysis is presented in Chapter 3.

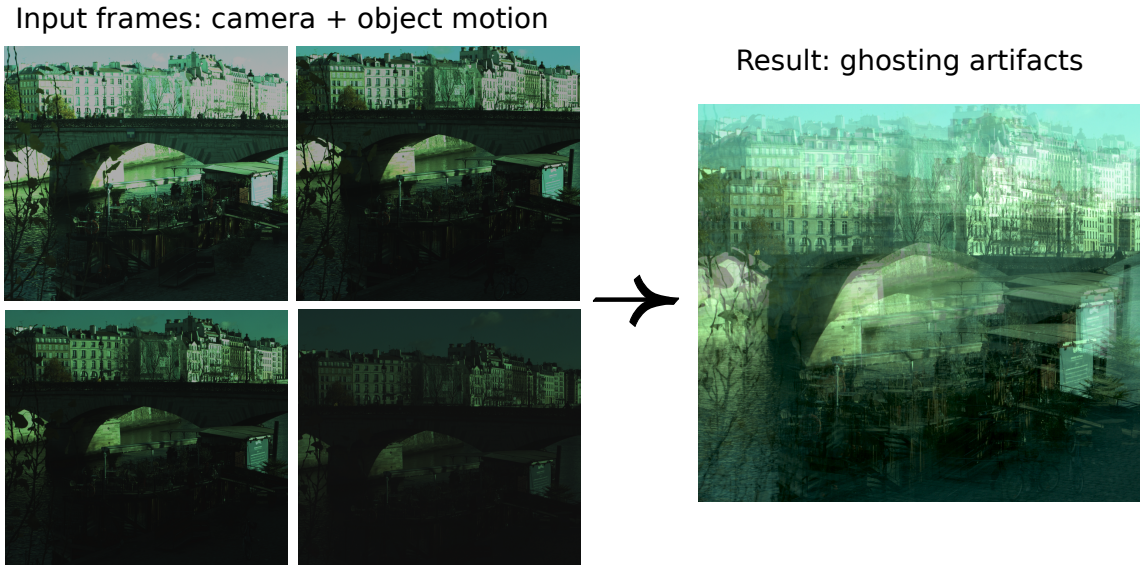


Figure 1.5: **Left:** Input misaligned images (these images are the result of demosaicking raw images, white balance has not been performed). **Right:** Result of the combination of the input frames without previous corregistration. Ghosting artifacts appear due to image misalignment.

Combined HDR and super-resolution The frequency content of a scene, as well as its dynamic range, is truncated when we take a single shot of the scene using a regular digital camera. Super-resolution (SR) techniques aim at recovering the lost high frequency information by combining multiple images captured with different sampling grids due to camera motion. Hence, the reconstruction of the dynamic range and that of the frequency content can be combined into a single problem and treated jointly.

From a collaboration with Yann Traonmilin, we propose an acquisition strategy that, if the affine motion hypothesis holds and sufficiently long exposure time is available, guarantees the recovery of the real high dynamic range and high frequency information of the original scene.

1.3 HDR imaging for dynamic scenes and hand-held camera

The study of the static case, despite being necessary to fully understand the task of HDR image generation, is often not directly applicable in practice. Indeed, most scenes present moving objects and a tripod is rarely available to keep the camera at a fixed position. The dynamic case brings very challenging problems to the generation of HDR images, as it does to all multi-imaging techniques. Global misalignments due to camera motion and local misalignments due to object motion make it impossible to directly combine the input images according to (1.2). Figure 1.5 shows an example of the result obtained when using (1.2) in the dynamic case. Ghosting artifacts appear all over the image giving a completely unusable result.

Most HDR techniques for dynamic scenes, first use some kind of global alignment

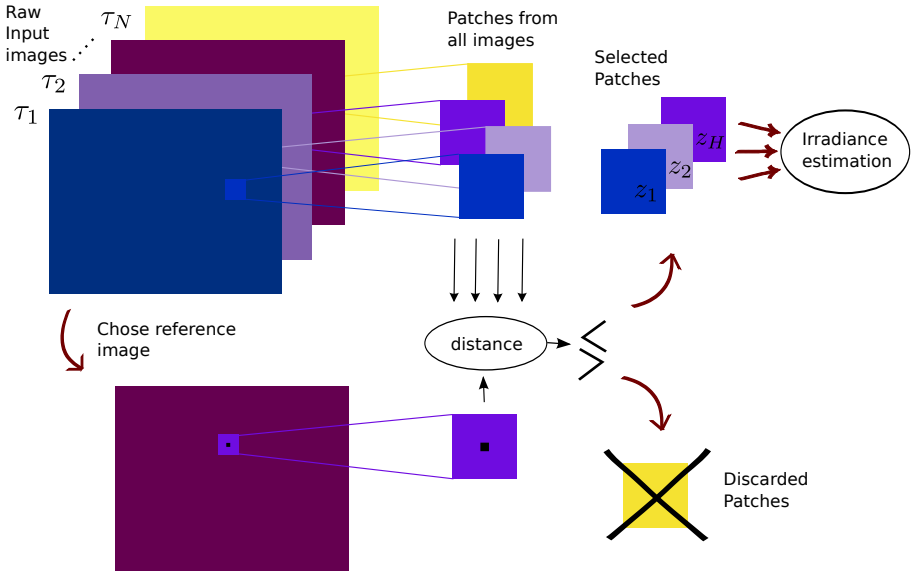


Figure 1.6: Diagram of the proposed method for multi-imaging HDR for dynamic scenes captured with a hand-held camera.

method to correct camera motion and a de-ghosting technique to correct object motion [Srikantha and Sidibé 2012]. Moreover, the pixel-wise estimation proposed in (1.2) does not take advantage of the redundant information existing in the input images, which has been proven very useful to improve image quality in various patch-based denoising techniques [Buades et al. 2008; Boracchi and Foi 2008].

In this work, we tackle these problems by taking advantage of the self-similarity existing in the input frames through a patch based approach. Two quite different approaches are proposed. The first one is a multi-image approach making use of a standard digital camera, as presented in the previous sections. The other is a single-shot acquisition strategy, based on the idea of spatially varying pixel exposures (SVE) introduced by Nayar and Mitsunaga [2000].

Non local HDR We first propose a multi-image approach that we call non local HDR. We address the previously mentioned problems by exploiting the self-similarity existing in each input frame, as well as between frames, using a patch based strategy.

Figure 1.6 shows a diagram of the proposed method. First, a reference frame is chosen among the T input frames captured with exposure times $\tau_1, \dots, \tau_T$. For each pixel p in the reference frame, its set of similar patches is found comparing a neighborhood of p to all patches contained in a search window centered at p in all the input frames. The patch distance is equal to the L^2 distance normalized by the noise variance of the patches, which is spatially variable and defined according to Model (1.1). If the patch distance to p is below a threshold, the patch is kept, otherwise it is discarded.

As it will be shown in more detail in Chapter 5, for HDR imaging, having an accurate model for the pixel's noise variance is crucial in order to correctly threshold the patch distance. Because images are captured with very different exposure times, they have very

Figure 1.7: **Real data.** Dynamic scene (pedestrians in the bridge and people next to the boat) acquired using a hand-held camera. **Left:** Tone mapped irradiance estimation using the proposed non local approach. No ghosting artifacts appear. **Right first row:** Input images (JPEG version). **Right second row:** Extracts of the the normalized reference image. **Right third row:** Extracts of the results by Sen et al. Sen et al. [2012]. **Right fourth row:** Extracts of the results by the proposed non local approach. The result obtained by the proposed approach is far less noisy than the one by Sen et al.. Please see the electronic copy for better color and details reproduction.

different noise levels, and a fixed threshold cannot correctly group very bright and very dark patches at the same time.

The central pixels of the set of similar patches are assumed to share the same underlying irradiance C_p. Therefore, because we showed that the MLE is nearly optimal for the irradiance estimation from T independent samples following the same distribution [Aguerrebere et al. 2014b] (Chapter 3), the irradiance at pixel p is computed as the MLE from the central pixels of the similar patches.

The usage of a reference image and a patch similarity to restore this reference image avoids the need of tedious global image registration and object motion detection. Indeed, patch based denoising techniques have been applied in the multi-image case and have been proven robust to image misalignments resulting from a hand-held camera as well as to object motion [Buades et al. 2005b]. Furthermore, combining information from multiple pixels gives a considerably less noisy result than current state-of-the-art HDR methods.

Figure 1.7 shows an example of the results obtained with the proposed patch-based approach. A comparison is made with the state-of-the-art method in HDR imaging for dynamic scenes proposed by Sen et al. [2012]. The result is also compared with the input reference image in order to show the original noise level. The denoising ability of

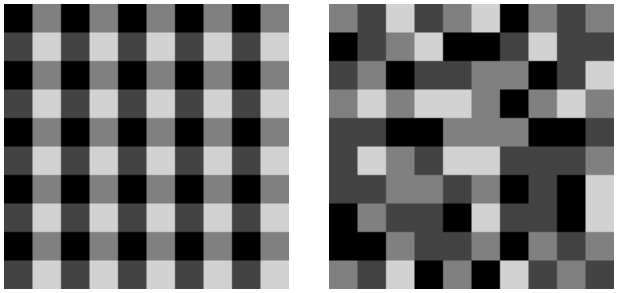


Figure 1.8: Regular (left) and non-regular (right) optical masks for an example of 4 different filters.

the proposed approach can be verified in the extracted details. Moreover, since no image alignment nor motion correction is needed, the result is completely free of ghosting artifacts.

Single-shot HDR Nayar and Mitsunaga [2000] were to our knowledge the first to introduce an acquisition strategy to create HDR images from a single shot. The idea is to place an optical filter, with spatially varying transmittance, adjacent to the conventional image sensor in order to control the amount of light reaching each pixel. Hence, a single shot captures an increased dynamic range compared to that of a conventional sensor. Figure 1.8 shows two examples of optical mask with four different transmittance levels. One corresponds to a regular sampling pattern [Nayar and Mitsunaga 2000] and the other to a random pattern [Schöberl et al. 2012a].

Being a single-imaging technique, this acquisition strategy leaves behind all the main drawbacks of multi-imaging techniques, such as global image mis-alignment and ghosting artifacts due to object motion. Another problem encountered by multi-imaging techniques based on a reference image, as the non local HDR estimation presented above, is the need to inpaint large unknown regions caused by saturation or under-exposure. SVE acquisition, which avoids having large regions to inpaint, overcomes this problem.

However, new challenges appear. Unlike the multi-imaging techniques, where all the pixels are correctly exposed in at least one of the input frames, in the single-shot acquisition the brighter (darker) regions will become saturated (under-exposed) and need to be restored somehow. Moreover, because less information is available in a single-shot method, the noise problem is more dramatic, mainly in the darker regions, and definitely needs to be addressed.

In this work, we propose to exploit the image self-similarity and the Gaussian mixture models, which have been proven accurate at representing image patches [Yu et al. 2012; Zoran and Weiss 2011], in order to restore the unknown pixels (saturated or under-exposed) and denoise the correctly exposed ones. A Bayesian general restoration method, proposed in this work and presented in Chapter 6, is applied to reconstruct the dynamic range content of the scene from a single image captured using SVE. As can be seen for example in Figure 1.9, the proposed method manages to simultaneously reconstruct the missing pixels and denoise the correctly exposed ones.

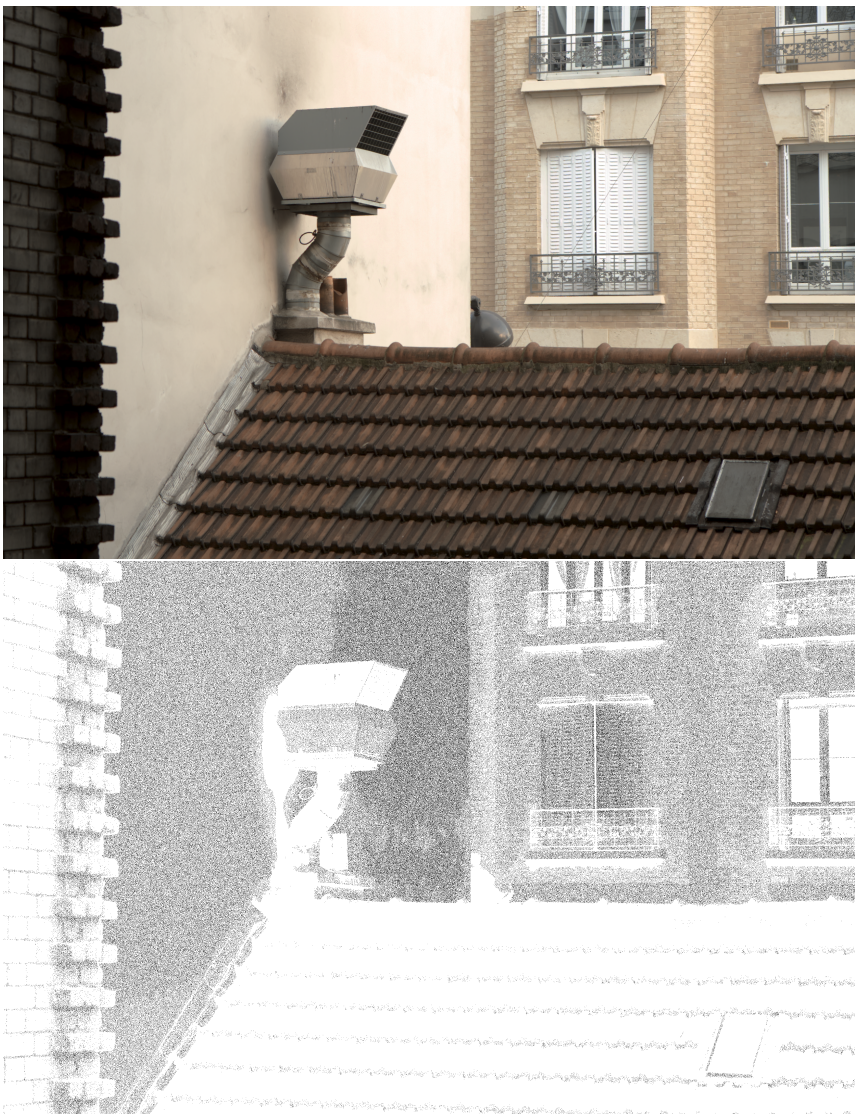


Figure 1.9: **Real data.** Tone mapped version of the HDR image obtained by the proposed approach (11.4 stops) and its corresponding mask of missing pixels (black). Please see the digital copy for a more accurate visualization.

1.4 A general image restoration method

The restoration method originally developed to generate HDR images from a single shot acquired using SVE, introduced in the previous section, is actually a general restoration method that can be used in several applications beyond HDR imaging. More precisely, it can be used in any application that requires the simultaneous reconstruction of random missing pixels and the denoising of existing ones.

Many examples can be found where this kind of degradation appears. For example the combined additive and impulsive noise present in raw images, the limited resolution and dynamic range as mentioned in Section 1.2, the noise and missing pixels as in the single SVE image HDR acquisition strategy introduced in the previous section. It can also applied, of course, to treat each problem separately, for the inpainting of random missing pixels or the classical denoising of additive noise.

Many of the recent state-of-the-art restoration methods tackle this kind of problems using patch-based approaches. Some of them rely on patch models and a Bayesian reconstruction method to restore degraded patches. In particular, Gaussian mixture models (GMM) have been proven successful to model natural image patches and the Bayesian framework has been proven adequate to exploit this patch prior and restore highly degraded patches.

In this work, we propose a new restoration approach that aims at simultaneously denoising and restoring missing pixels. It is inspired by the combination of two powerful patch-based restoration methods: the general framework for the resolution of inverse problems proposed by Yu et al. [2012], called Piece-wise Linear Estimators (PLE) and the non-local Bayesian based denoising algorithm (NLB) proposed by Lebrun et al. [2013]. We combine the strengths of both methods into a single algorithm. On the one hand, the ability of PLE to restore missing pixels, and on the other hand the denoising capacity of NLB.

The proposed method makes use of a Gaussian prior for image patches, whose mean μ and covariance matrix $\mathbf{\Sigma}$ are estimated locally from similar patches. In Bayesian statistics, μ and $\mathbf{\Sigma}$ are known as hyperparameters, since they are the parameters of a prior distribution, while a prior on them is called an hyperprior. In order to deal with missing pixels, which is not possible in NLB, we rely on an hyperprior for μ and $\mathbf{\Sigma}$. This hyperprior is used to compute a MAP estimator for the Gaussian parameters μ and $\mathbf{\Sigma}$. Then, the patch restoration is performed by computing the MAP estimator of the patches with the locally estimated Gaussian prior.

The proposed approach has been tested in a series of experiments presenting different degradation scenarios: random missing pixels with different masking rates, additive Gaussian noise with different noise levels and a combination of both. Results show that the restoration capability of the proposed approach outperforms that of PLE, both for plain interpolation and in the case of combined interpolation and denoising. The advantage of the proposed method is specially clear for high masking rates. On the other hand, the denoising power is shown to be very similar to that of NLB, which was expected since in the case of denoising only (no missing pixels) the estimation performed by both methods is quite close.

Outline of the document

This document is organized as follows:

Chapter 2: Introduces the camera acquisition model used throughout the work.

Chapter 3: Presents the computation of the performance bounds, the inclusion of saturation information and the study of the influence of the uncertainty in the camera parameters on the irradiance estimation.

Chapter 4: Presents the proposed HDR-SR acquisition strategy.

Chapter 5: Describes the proposed multi-image approach for HDR imaging with dynamic scenes and hand-held camera.

Chapter 6: Presents the proposed general Bayesian patch-based reconstruction method with an application to HDR imaging from a single shot captured with spatially varying pixel exposures.

The bibliographic analysis for each subject is performed separately in the corresponding chapters. Hence, the one corresponding to the static case can be found in Chapter 3, the one for the dynamic case in Chapter 5 and the one for the single-image HDR in Chapter 6.

Publications

The following articles were published during this thesis

- Y. Traonmilin and C. Aguerrebere, *Simultaneous High Dynamic Range and Super-Resolution Imaging Without Regularization*, Accepted at SIAM Journal on Imaging Sciences in April 2014.
- C. Aguerrebere, J. Delon, Y. Gousseau and P. Musé, *Best Algorithms for HDR Image Generation. A Study of Performance Bounds*, SIAM Journal on Imaging Sciences. vol. 7, no. 1, pp. 1-34, 2014, doi 10.1137/120891952
- C. Aguerrebere, J. Delon, Y. Gousseau and P. Musé, *Simultaneous HDR image reconstruction and denoising for dynamic scenes*, IEEE International Conference on Computational Photography (ICCP), 2013, pages 1 - 11, doi: 10.1109/ICCPhot.2013.65283009.
- C. Aguerrebere, A. Almansa, J. Delon, Y. Gousseau and P. Musé, *Single Shot High Dynamic Range Imaging Using Piecewise Linear Estimators*, IEEE International Conference on Computational Photography (ICCP), 2014.
- C. Aguerrebere, Y. Gousseau and G. Tartavel, *Exemplar-based Texture Synthesis: the Efros-Leung Algorithm*, Image Processing On Line, vol. 3, pp. 223-241, 2013, doi 10.5201/ipol.2013.59

Poster presentations

- C. Aguerrebere, J. Delon, Y. Gousseau and P. Musé, *Algorithmes optimaux pour la génération d´images HDR: Une etude des bornes de performance*. Colloque Gretsi, Brest, September 2013.

- C. Aguerrebere, A. Almansa, J. Delon, Y. Gousseau and P. Musé, *High dynamic range imaging from a single shot*, SIAM Conference on Imaging Science, Hong Kong, May 2014.

2 Camera acquisition model

The accurate modeling of the acquisition process in digital cameras is of great interest for a wide variety of domains concerning the use of digital images. In particular, it is of great utility in image processing, computational photography or computer vision applications. For instance, the statistical characterization of image data allows to develop denoising techniques suited to particular noise types, which may outperform general techniques.

In this chapter we present a detailed analysis of the digital image acquisition process which allows us to introduce a statistical model of the sensor raw data. The accuracy of this model is essential to its posterior utility. It is thus crucial to take into account all the main noise and uncertainty sources in the proposed model. This model is fundamental for the work presented hereafter since most of the performed analysis and proposed methods are based on it.

Several articles present and make use of statistical models of the sensor raw data. In HDR imaging, we can cite the works by [Tsin et al. 2001; Robertson et al. 2003; Kirk and Andersen 2006; Granados et al. 2010; Foi et al. 2008]. Different levels of complexity can be found among them. In particular, a model similar to the one presented in this chapter can be found in [Kirk and Andersen 2006; Granados et al. 2010; Foi et al. 2008]. Nevertheless, to the best of our knowledge, none of these articles present a detailed explanation of the physical origin of each noise source and the corresponding justification of the statistical model associated to each one of them. This rigorous analysis enables us to prioritize the different noise sources and obtain a simplified model which is both usable and realistic.

2.1 Acquisition of digital images

Two technologies are used for camera sensors: charge-coupled devices (CCD) and complementary metal–oxide–semiconductors (CMOS). Even if the operation principles of both sensors differ, a very similar acquisition model can be proposed for both of them, illustrated by a simplified diagram in Figure 2.1. In short, CCDs and CMOS both transform incoming light photons into voltage output values. More precisely, these sensors are silicon-based integrated circuits including a dense matrix of photo-diodes that first convert light photons into electronic charge [Theuwissen 1996; Brouk et al. 2008]. Light photons interact with the silicon atoms generating electrons that are stored in a potential well. When the potential well is full, the pixel saturates, and no further electrons are

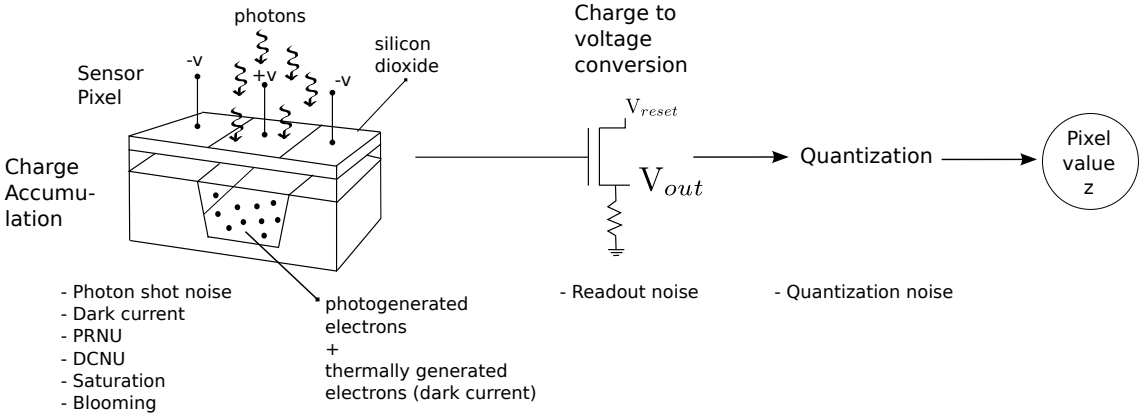


Figure 2.1: Simplified diagram of the main stages of the acquisition process and the principal noise sources at each stage.

stored [1].

In the case of CCDs, the accumulated charge may then be efficiently transferred from one potential well to another across the chip, until reaching an output amplifier where the charge is converted to a voltage output value. This voltage is then quantified to give the corresponding pixel value. For the CMOS technology, the impinging photons are also accumulated in the photo-diodes. However, unlike CCDs, CMOS pixels have conversion electronics to perform the charge to voltage conversion at each location. This extra circuitry increases noise and generates extra fixed pattern noise sources compared to CCDs [Brouk et al. 2008].

The main uncertainty sources at each stage of the acquisition process are described in more details in the following paragraphs, and listed in Figure 2.1. We divide them into two categories: random noise sources, and spatial non-uniformity sources.

2.1.1 Random noise sources

Two physical phenomena are responsible for the random noise generation during the camera acquisition process: the discrete nature of light, which is the cause of the photon shot noise, and thermal agitation, which explains the random generation of electrons inside the sensor when the temperature increases.

Photon shot noise The number of photons C_i^p impinging the photo-diode p during a given exposure time τ_i follows a Poisson distribution, with expected value $C_p\tau_i$, where C_p is the radiance level (in photons per unit-of-time) reaching the photo-diode. If we suppose that an electron is generated for each absorbed photon (this depends on the photon energy, therefore on the considered wavelength), the number of electrons generated on the potential well is also Poisson distributed. In an ideal case with no other noise sources, the voltage measured at the sensor output should be proportional to the collected charge:

[1] In this case, additionally generated electrons may spill over the adjacent wells, resulting in what is called *blooming*. This phenomena, well known in astronomic photography, is mostly observed with very long exposures. We neglect it in this work.

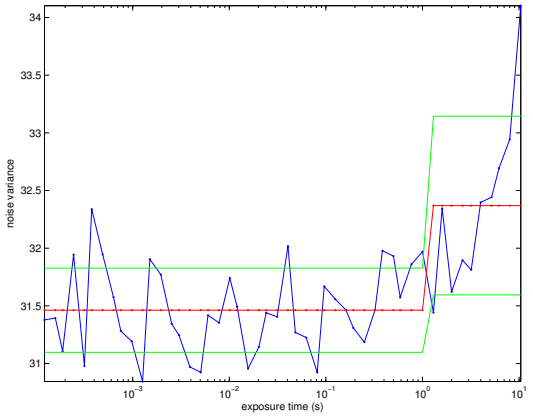


Figure 2.2: Noise variance vs. exposure time for dark frames (blue). The red "step" shows the mean noise variance for the exposures below 1 second and then for those above 1 second. The band delimited by the green lines represents one standard deviation from the mean value. Notice that the noise increases for exposures above 1 second.

$V = g_{cv} C_p \tau_i$, where $C_p \tau_i$ is the number of absorbed electrons, and where g_{cv} is the equivalent capacitance of the photo-diode.

Dark current Some of the electrons accumulated in the potential well do not come from the photon to electron conversion but result from thermal generation. These electrons are known as *dark current*, since they are present and will be sensed even in the absence of light. Dark currents can be generated at different locations in the sensor and are related to irregularities in the fundamental crystal structure of the silicon, e.g. metal impurities (gold, copper, iron, nickel, cobalt) and crystal defects (silicon interstitials, oxygen precipitates, stacking faults, dislocations) [Theuwissen 1996]. For an electron to contribute to the dark current it must be thermally generated but also manage to reach the potential well. This last event happens independently for each electron. As a consequence, it can be shown that the number of electrons D^p thermally generated and reaching the potential well p is well modeled by a Poisson distribution with expected value D^p [Theuwissen 1996], depending on the temperature and exposure time. This noise is generally referred to as dark *current shot noise* or *dark shot noise*. In this paper, in order to make explicit the dependence on the exposure time τ_i, we name this dark shot noise D_i^p.

It has been stated that dark currents can be neglected for exposure times under 1 second [Martinec b]. The following experiment was conducted to verify this result for the Canon 7D camera. Dark frames (frames acquired in a dark room with a camera without lens and with the cap on) were acquired for exposure times in the range $1/8000$ to 10 seconds and ISO set to 100. The noise variance for each frame was computed as the variance from all pixels. This variance includes the readout noise variance and the dark current shot noise variance. Figure 2.2 shows the obtained results. The noise variance is nearly constant up to 1 second and then increases with the exposure time for exposures above 1 second.

The dark current expected value increases with the exposure time. Since this behavior is not observed for the variance values between 1/8000 and 1 second, we conclude that

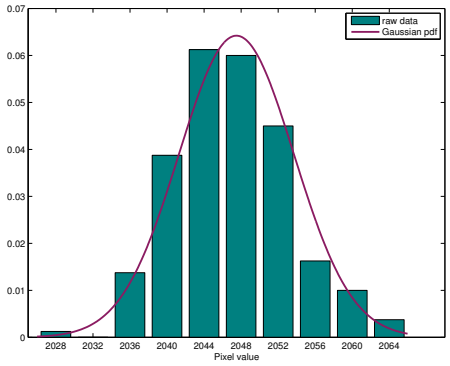


Figure 2.3: **Readout noise**. **Green**: histogram of the raw values taken by 200 realizations of one pixel acquired with virtually no light (realizations are obtained from 200 bias frames acquired with a Canon 7D camera set to its shortest exposure 1/8192.0 s). **Violet**: Gaussian probability density function with mean and variance computed from the raw data. The distribution of the raw pixels is accurately approximated by a Gaussian distribution.

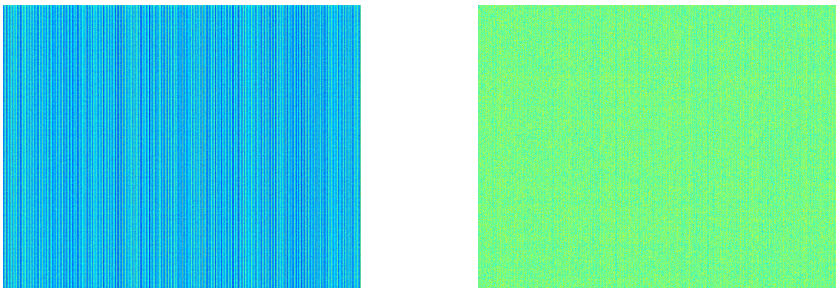
the dark current variance is masked by the readout constant variance and is therefore negligible with respect to it for exposure times below 1 second.

Readout noise In the readout stage of the acquisition process a voltage value is read for each pixel. This voltage is read as a potential difference from a reference level which represents the absence of light. Thermal noise N_{reset}, inherent to the readout circuitry, affects the output values. In the literature, it is widely admitted that this noise is accurately modeled as Gaussian distributed [Mancini 2002]. It is also known as *reset noise*, in reference to the reference voltage, commonly named reset voltage.

Notice that modeling the noise source as Gaussian distributed means that pixels may take negative values. In some cameras, the reference voltage is assigned a large enough value in the AD conversion so that voltage values below the reference are assigned positive pixel values. For this reason, the raw data for an image taken with the cap on will give pixel values close to the offset value (e.g. 2048 for the 14 bits Canon 7D). Alike the raw pixels, after subtracting the offset, the inverse of the camera response $f^{-1}(\mathrm{z})$ may take negative values.

The readout noise N_{out} includes also the remaining circuitry noise sources between the photoreceptor and the AD circuitry. They are all thermally generated and thus modeled as Gaussian noise. Some other minor sources include frequency dependent noise (flicker noise) but we wont consider them in this analysis.

Figure 2.3 shows the histogram of the raw values taken by 200 realizations of one pixel. These realizations are obtained from 200 bias frames (images acquired with virtually no light, i.e, in a dark room with a camera without lens and with the cap on using the shortest possible exposure) acquired with a Canon 7D camera set to its shortest exposure (1/8192.0 s). The camera acquires virtually no light, thus the pixel value captures the readout noise in that pixel (the dark current can be neglected as shown in the previous experiment). The Gaussian probability density function, with mean and variance computed from the raw data, is superposed for comparison. The Gaussian distribution accurately approximates



(a) Average bias frame. (b) Bias frame.

Figure 2.4: **Column noise**. **Left**: Average of 200 bias frames acquired with a CMOS sensor. **Right**: Bias frame acquired with a CMOS sensor.

the readout noise distribution. Moreover, this experience shows the presence of the offset value previously mentioned. Even if the images are acquired with no light, the mean pixel value is not zero but 2048.

2.1.2 Spatial non-uniformity sources

Besides random noise sources, several uncertainty factors, all related to the spatial non-uniformity of the sensor, should be taken into account in the acquisition model.

Fixed pattern noise sources

Photo-response non-uniformity (PRNU) The PRNU describes the differences in pixel responses to uniform light sources. Different pixels will not produce the same number of electrons from the same number of impacting photons. We assume one electron is generated per absorbed photon, but not all the impinging photons will be absorbed in the photo-diode. This is caused by variations in the pixel geometry, substrate material and micro-lenses [Irie et al. 2008]. The effect of PRNU is proportional to illumination and is prominent under high illumination levels.

The fact that a photon can be absorbed or not in the photo-diode is a binomial selection of the Poisson process of impinging photons. Hence the PRNU can be modeled as a multiplicative factor a_p applied to the parameter of the Poisson variable C_i^p.

Dark-current non-uniformity (DCNU) The DCNU represents the variations in dark current generation rates from pixel to pixel. This variation is intrinsic to the material characteristics of the sensor cells and causes variations in the expected value of the dark current from pixel to pixel. As the PRNU, the DCNU can be modeled as a multiplicative factor d_p applied to the parameter of the Poisson variable D_i^p.

CCD specific sources

Transfer efficiency After charge is collected at each pixel, the CCD must transfer it to the output amplifier for readout. The transfer efficiency of a real CCD sensors is less

Signal Dependent	Temperature Dependent	Exp. Time Dependent
Photon shot noise	Thermal noise	Photon shot noise
PRNU	Dark current	Dark current shot noise
		Thermal noise

Table 2.1: Classification of the noise sources according to their dependence on signal, temperature and exposure time.

than 1. Charge that is not correctly transferred is either lost or deferred to other transfers, affecting other pixels count values. Current buried-channel CCD transfer efficiency is above 0.99999 [Healey and Kondepudy 1994] thus it will not be taken into account in the acquisition model.

CMOS specific sources

Column noise The readout for CMOS sensors is performed line by line. At a given time, all columns of one line are readout through the output column amplifiers. Differences from one column amplifier to another introduce a column fixed pattern. Because the human eye is adapted to perceive patterns, column noise may be quite disturbing even if its contribution to the total noise is less significant than that of white noise [Martinec a].

Figure 2.4a shows the average of 200 bias frames acquired with a CMOS sensor (Canon 7D set to exposure time 1/8192.0 s). The column pattern on the readout noise is clearly visible. Figure 2.4b shows an example of one bias frame. The column pattern is not so evident from just one frame, but a subtle column pattern is still noticeable.

Another interesting classification of the noise sources relies on their dependence on signal, temperature and exposure time. Table 2.1 shows such a classification of the different noise sources.

2.1.3 Quantization noise

A last source of noise in the acquisition process takes place during the conversion of the analog voltage measures into digital quantized values or data numbers (DN). When the signal variation is much larger than 1 DN [2], the quantization noise can be modeled as additive and uniformly distributed. In that case, it is usually negligible compared to the readout noise [Healey and Kondepudy 1994]. This being even more remarkable for modern cameras, which can easily have 12 or 14 bits for quantization.

The following experiment was conducted to verify the previous statement. The variance of a bias frame gives the variance caused by the readout process including the quantization noise. Several tests were performed with a Canon 7D and a Canon 400D, acquiring bias frames with different ISO settings. The variance values obtained are: Canon 400D, ISO 100 var = 2.5, ISO 400 var = 6.3, ISO 800 var = 17.2. Similar values are found in [Granados et al. 2010] for a Canon 5D set to ISO 400 (var = 6.5) and a Canon PowerShot S5 set to

[2] This is the case for non low light conditions since under the Poisson model, the irradiance mean and variance are equal so a low signal implies low variance.

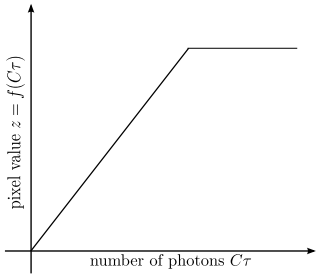


Figure 2.5: Camera response function for the linear + saturation model (raw data).

ISO 400 (var = 18). As previously stated, except in low light conditions, the quantization noise can be modeled as uniformly distributed with variance equal to 1/12. Hence, in all of the tested configurations, the contribution of the quantization noise to the total readout variance is negligible (in the worst case 2.5 compared to 1/12). Moreover, the variance of the quantization noise in the worst case scenario (regardless of the illumination level) is 1/4, the worst possible error being 1/2. Thus, the contribution of the quantization noise to the total readout variance can also be neglected in low light conditions for the tested configurations (1/4 compared to 2.5). However, the margin is not so large in that case and counterexamples are certainly possible.

Given the recent advances in digital images acquisition techniques, and the corresponding decrease on the readout noise values, it may be interesting to include the quantization noise effects in order to develop a more precise camera acquisition model.

It is known that even if the power of quantization noise is negligible with respect to the other noise sources, its structured nature (it is not white) may make it noticeable after non linear post-processing. Nevertheless, when considering a noise model for raw data only, i.e., before non linear post-processing, the hypothesis of negligible quantization noise, except in low light conditions, remains valid.

2.1.4 Acquisition model

Equation (2.1) proposes a simplified model including the previous noise sources. Pixels are modeled as independent and following the same model (the dependence on position p is avoided to simplify the notation):

$$\mathrm{Z}_i = f\Big(g_{out}\left[g_{cv}(\mathrm{C}_i + \mathrm{D}_i) + \mathrm{N}_{reset}\right] + \mathrm{N}_{out} + \mathrm{Q}\Big), \tag{2.1}$$

where Z_i is the pixel value, f is the camera response function, g_{out} is the camera gain, and Q is the uniformly distributed quantization error. The term

$$g_{out}\left[g_{cv}(\mathrm{C}_i + \mathrm{D}_i) + \mathrm{N}_{reset}\right] + \mathrm{N}_{out}$$

can be rewritten as the addition of a Poisson distributed random variable with expected value $\lambda_i = a\tau_i C + dD_i$, multiplied by the gain factor $g = g_{cv}g_{out}$, and a Gaussian distributed noise component $\mathrm{N}_R = g_{out}\mathrm{N}_{reset} + \mathrm{N}_{out}$ with mean μ_R and variance σ_R^2.

A similar model is presented by Foi et al. [2008], where they propose to model digital camera raw data as a mixed Poisson-Gaussian model. The difference between the models

is the inclusion of the gain g_{cv}, modeling charge to voltage conversion. This constant is not included in Foi et al. model, but the general idea remains the same.

The previous model is valid for both CCD and CMOS sensors. In the CCD case, the readout noise sources can be considered as identical for all pixels. Thus g, μ_R and σ_R^2 are spatially constant. On the contrary, in order to model column noise for CMOS sensors, different g, μ_R and σ_R^2 parameters should be considered for each column. In this work we decide not to include the column noise pattern in the proposed model and consider g, μ_R and σ_R^2 as spatially constant parameters.

For the values usually taken by $\lambda_i = a\tau_i C + dD_i$, the Poisson distribution can be correctly approximated by a Gaussian distribution with mean and variance equal to λ_i. Regarding the relative importance of each noise source, under low illumination conditions the primary noise source is the reset noise, while for high illumination the major noise source is the photon shot noise [Healey and Kondepudy 1994]. The dark currents can be neglected for exposure times below 1 second [Martinec b; Aguerrebere et al. 2012] and, except in low illumination conditions, the quantization noise can be neglected compared to the readout noise [Healey and Kondepudy 1994; Aguerrebere et al. 2012].

Simplified model As a consequence of the previous statements, and assuming $\tau_i < 1$ second, we can assume that the variable $f^{-1}(\mathrm{Z}_i)$ follows a Gaussian distribution $\mathcal{N}(ga\tau_i C + \mu_R, g^2 a\tau_i C + \sigma_R^2)$. In the case of raw data, f is a linear function before attaining its saturation threshold (see Figure 2.5). Thus for non saturated samples the model becomes

$$\mathrm{Z}_i \sim \mathcal{N}(ga\tau_i C + \mu_R, g^2 a\tau_i C + \sigma_R^2). \tag{2.2}$$

This model will be used as a starting point in the analysis and methods proposed along this work as presented in the following chapters.

2.2 Camera calibration procedure

In this section we describe the camera calibration procedure followed in this work in order to estimate the camera parameters. This procedure is based on the work by Granados et al. [2010].

Readout noise mean and variance The readout noise mean and variance are computed from a bias frame. According to Model (2.2), where the readout noise mean and variance do not depend on the pixel position, these can be obtained computing the empirical mean and variance from all the pixel values of the bias frame,

$$\hat{\mu}_R = \frac{1}{N}\sum_{j=1}^{N} \mathrm{b}_j \qquad \hat{\sigma}_R^2 = \frac{1}{N-1}\sum_{j=1}^{N} (\mathrm{b}_j - \mu_R)^2, \tag{2.3}$$

where b_j, $j = 1, \ldots, N$ are the pixel values of the bias frame.

This estimation neglects the variance of the dark current and the quantization noise with respect to that of the readout noise.

Gain In order to estimate the camera gain we first note that, from Model (2.2), the mean and variance of any pixel Z verify

$$\mu_{\mathrm{Z}} = ga\tau C + \mu_R \tag{2.4}$$

$$\sigma^2_{\mathrm{Z}} = g^2 a\tau C + \sigma^2_R, \tag{2.5}$$

thus we have

$$g = \frac{\mu_{\mathrm{Z}} - \sigma^2_R}{\sigma^2_{\mathrm{Z}} - \mu_R}. \tag{2.6}$$

Flat fields (i.e., images captured illuminating the sensor with a spatially uniform, narrow band light source) are thus used to estimate the pixels mean and variance and compute the corresponding g value. The pixels mean is computed according to

$$\hat{\mu}_{\mathrm{Z}} = \frac{1}{N}\sum_{j=1}^{N} \mathrm{z}_j, \tag{2.7}$$

where z_j, $j = 1, \ldots, N$ are the pixel values of a flat field. The pixels variance is computed from the difference of two flat fields in order to compensate for the variance introduced by the PRNU effect

$$\hat{\sigma}^2_{\mathrm{Z}} = \frac{1}{2(N-1)}\sum_{j=1}^{N} (\mathrm{z}^1_j - \mathrm{z}^2_j)^2, \tag{2.8}$$

where z^1_j and z^2_j are the corresponding pixels of the two flat fields. Finally we have,

$$\hat{g} = \frac{\frac{1}{2(N-1)}\sum_{j=1}^{N} (\mathrm{z}^1_j - \mathrm{z}^2_j)^2 - \sigma^2_R}{\frac{1}{N}\sum_{j=1}^{N} \mathrm{z}_j - \mu_R}. \tag{2.9}$$

In practice, the creation of a flat field is not a trivial task and can highly complicate the calibration procedure. An alternative to this is to compute the gain factor from image sub-regions that correspond to flat regions. Instead of computing the gain factor from all image pixels, we use all the pixels in the image sub-regions that correspond to the same illumination level. This is possible since the gain factor is a global parameter (i.e. the same for all pixel values), and its value should remain the same either computed from the complete flat frame or from a corresponding sub-region. In order to have a more robust estimate, the average of several estimates from flat sub-regions should be considered.

Photo response non uniformity Flat fields can also be used to estimate the spatially varying factors of the photo response non uniformity. From Model (2.2) we have

$$\mu_{\mathrm{Z}_j} = ga_j\tau C + \mu_R, \tag{2.10}$$

where Z_j denotes the j-th pixel location of the flat field. The mean μ_{Z_j} can be estimated from M flat fields as

$$\hat{\mu}_{\mathrm{Z}_j} = \frac{1}{M}\sum_{m=1}^{M} \mathrm{z}^m_j, \tag{2.11}$$

where z_j^m denotes the j-th pixel location of the m-th flat field. Then the PRNU factor at location j can be estimated as

$$\hat{a}_j = \frac{\hat{\mu}_{Z_j} - \mu_R}{\frac{1}{N}\sum_{h=1}^{N} \hat{\mu}_{Z_h} - \mu_R}. \tag{2.12}$$

Saturation threshold Most saturated samples are clearly identifiable since they are assigned the maximum output pixel value. However, some saturated samples may be assigned a slightly inferior value due to readout noise. For this reason, the saturation threshold should be set to a percentage (98% in our tests) of the maximum output pixel value.

3 HDR imaging for static scenes and static camera

As explained in the introduction, high dynamic range (HDR) imaging is the field of imaging that seeks to accurately capture and represent scenes with the largest possible irradiance range. Due to technological and physical limitations of current optical sensors, nowadays the most common way to reach high irradiance dynamic ranges is by combining multiple low dynamic range photographs, acquired with different exposure times $\tau_1, \tau_2, \ldots, \tau_T$[1]. Indeed, for a given irradiance C and exposure time τ_i, the corresponding pixel value is a function of the received luminous energy $\tau_i C$. Hence, using different exposure times allows to sample the camera response function at different operating points thus avoiding saturation for at least some of the exposures and keeping details in both dark and bright regions. This acquisition process is called exposure bracketing. Of course, the more photographs spanning the whole range of exposure times, the better the result. However, for practical limitations, exposure bracketing in real situations can rarely exceed a few snapshots, say 2 to 6.

In this chapter we concentrate on the problem of estimating the irradiance map, that is the irradiance reaching each pixel, from a reduced number of photographs captured in controlled conditions (static scene and static camera) with a given set of exposure times. More precisely, we conduct a thorough analysis of the HDR image estimation problem, in order to establish its performance limits and to determine if current estimation techniques are close to these limits. Part of this study also consists in quantifying how these limits are affected when samples saturate, and when uncertainties are introduced in the camera parameters calibration process. The analysis presented in this chapter is conducted under the following hypotheses:

1. The camera response function is assumed linear. The estimation is done from raw samples for which, ignoring noise sources, the camera response function is linear.

2. The photographs are perfectly co-registered and possible radiometric changes have been compensated.

3. The ISO setting is assumed to be fixed for all images, thus the gain factor is constant.

[1] An alternative method, using a single image acquired with spatially varying pixel exposures, is presented in Chapter 6.

4. The considered estimators are unbiased or nearly unbiased and the computed bounds set the performance limits for unbiased estimators. This hypothesis is motivated by the fact that most of the existing estimators, among them the best performing [Granados et al. 2010] and most well known ones [Kirk and Andersen 2006; Mann and Picard 1995; Debevec and Malik 1997; Mitsunaga and Nayar 1999; Reinhard et al. 2005], are nearly unbiased. Biased estimators are sometimes preferred since they may achieve a lower mean squared error (MSE) than unbiased ones. Nevertheless, we do not study the performance bound for biased estimators since it depends on the bias and must therefore be stated for each bias type. Notwithstanding, notice that the methodology here conducted can be easily extended to a given type of biased estimators by adapting the computation of the bound including the corresponding correction term (first derivative of the bias) and repeating the experimental stage.

Assuming that these hypotheses are satisfied, we analyze in this chapter the following aspects:

1. We present a study of the theoretical performance bounds of the HDR estimation problem. More precisely, given a small number of samples per pixel (say 2 to 6), we determine the lowest MSE that can be attained by combining these samples. One question we address is why the MLE outperforms the other estimators proposed in the literature, and how far it is from the optimal MSE. The optimality of the MLE is far from obvious in such a non asymptotic case. We show, however, that there is not much room for improvement.

2. Surprisingly, all the methods proposed in the literature discard saturated samples. Nevertheless, saturated samples certainly carry some useful information. For instance the exposure time at which they have saturated. As a second contribution, we study the usefulness of this information, and we present a way to incorporate it in the whole estimation process. Not surprisingly, the information contained in saturated samples turns out to improve the irradiance estimation.

Another aspect treated in this work deals with a very significant problem that does not seem to have received much attention in the literature: the sensitivity of irradiance estimation to uncertainties in the camera parameters. This is a crucial question, since these parameters are obtained through a calibration process, which is of course not error-free. Through an experimental analysis, we show that irradiance can only be accurately estimated if the camera parameters are very carefully calibrated. Moreover, a very careful calibration is also found to be critical in order to obtain a reliable ground-truth for evaluations with real data. Even if this part of the study is conducted with a limited number of experiments, the interesting preliminary results open a path to explore this topic.

The work presented in this chapter has been published in the SIAM Journal on Imaging Sciences [Aguerrebere et al. 2014b].

The chapter is organized as follows. In Section 3.1 we present a summary of existing HDR methods for the case of static scene and static camera. In Section 3.2, we derive performance bounds for the HDR estimation problem, and we compare some state-of-the-art

estimators against the bound. The use of saturated samples in the irradiance estimation problem is analyzed in Section 3.3. In Section 3.4 we present a sensitivity analysis of the estimation problem with respect to uncertainties in camera parameters. The main conclusions of this chapter are summarized in Section 3.5.

3.1 Previous work

To the best of our knowledge, the first HDR imaging technique in the framework of digital photography, based on exposure bracketing, was proposed in 1995 by Mann and Picard. The method assumes that the camera radiometric response function f mapping luminous energy to pixel values has been previously calibrated. Then, if z_i^p denotes the image value at pixel p for the exposure time τ_i, the irradiance estimate at that position $\hat{C}_p$ is computed from exposure times $\tau_1, \ldots, \tau_T$ as

$$\hat{C}_p = \frac{\sum_{i=1}^{T} w_i^p \frac{f^{-1}(z_i^p)}{\tau_i}}{\sum_{i=1}^{T} w_i^p}, \tag{3.1}$$

were w_i^p is the weight assigned to the exposure i for pixel p. In [Mann and Picard 1995], small weights are assigned to extreme pixel values – very low or close to saturation – based on the claim that the camera response function calibration is less accurate at these values.

This approach suffers from mainly two problems. The first one is that the weights are somehow arbitrary and not derived from a noise model of the pixel values. The second one is that the calibration of the camera response function is a problem on its own, and is prone to errors that are directly transferred to the irradiance estimator. The work of Mann and Picard represents a very important contribution, since it inspired several approaches based on exposure bracketing, whose main difference relies on the way photographs are combined, that is, on the choice of the weights w_i^p. See [Kirk and Andersen 2006; Granados et al. 2010] for an interesting review and comparison of these methods.

Irradiance estimation methods can be classified according to different criteria. One of them is whether the method assumes a linear [Robertson et al. 2003; Kirk and Andersen 2006; Granados et al. 2010] or a non-linear camera response function f [Mann and Picard 1995; Debevec and Malik 1997; Mitsunaga and Nayar 1999; Tsin et al. 2001; Reinhard et al. 2005]. The former are meant to be used with the camera raw data, i.e. the pixel values before any camera post-processing (demosaicking, white balance, etc). For raw data, the camera response function is linear since, ignoring noise sources, each pixel value is proportional to the number of photons reaching the corresponding sensor cell. The latter need to define a method to estimate the camera response function and its inverse.

In this work we will assume a linear camera response. With the currently available technology and storage capacity of cameras, it is entirely reasonable to assume that we have access to the raw data or that the processing can be done directly on the raw data inside the camera. For those methods like Mann and Picard's, which consider non-linear response functions, we will therefore consider the linear response counterpart based on raw data.

Another meaningful classification of irradiance estimation methods is based on the objective function that is optimized. Three main groups can be distinguished, whether the objective function is based on the camera response function [Mann and Picard 1995; Debevec and Malik 1997], the SNR [Mitsunaga and Nayar 1999; Reinhard et al. 2005] or the variance [Tsin et al. 2001; Robertson et al. 2003; Kirk and Andersen 2006; Granados et al. 2010; Hasinoff et al. 2010]. The methods in the first group [Mann and Picard 1995; Debevec and Malik 1997] propose to compute the weights based on the uncertainty of the sample values given by the camera response function. It is claimed that extreme pixel values are less accurate. Therefore lower weights are assigned to those values and mid-range values are prioritized. The SNR based methods [Mitsunaga and Nayar 1999; Reinhard et al. 2005] weigh samples according to their SNR, which is computed from the input samples.

Variance based approaches are among the most recent ones. An idea common to all of them is the use of a statistical model of the camera acquisition process. This model takes into account several noise sources allowing to improve the irradiance estimation. Tsin et al. [2001] were the first ones to propose this kind of approach. They characterize each pixel as a random variable whose distribution parameters depend on the unknown irradiance. Then they estimate the irradiance as an average weighted by the standard deviation of the samples. Robertson et al. [2003] propose a statistical model where the mean pixel value depends on the unknown irradiance. Then they compute the irradiance with a maximum likelihood estimator (MLE). The MLE is the average of the input samples weighted by the inverse of their variances. However, since they do not manage to estimate these variances they fix their values by an ad-hoc procedure.

More recently, a new variance-based approach was introduced by Kirk and Andersen [2006]. The main difference with the previous ones relies on the camera acquisition model. In this model, both the mean and the variance of the pixels depend on the unknown irradiance (because of the Poisson nature of the photon shot noise). They propose to estimate the irradiance also using a MLE. However, under this model a closed-form does not exist for the MLE. To overcome this limitation they ignore the variance dependence on the irradiance and find a closed-form for the MLE which depends on the variance. Then, they use the known dependence of the variance on the irradiance to compute the variance from the input samples.

The same line of work is followed by Granados et al. [2010] who propose a still more accurate model and solve the variance-irradiance dependence problem with an iterative method. Basically, at each iteration the weights are computed from the current irradiance value and the irradiance is updated according to the MLE (the one obtained ignoring the variance dependence on the irradiance).

In the imaging industry, a classical choice is the one obtained by setting $w_i^p = \tau_i$ in Equation (3.1). The estimator thus obtained is the minimum variance unbiased estimator when the input samples are assumed to follow a Poisson distribution with mean and variance depending on the irradiance parameter. As presented in Chapter 2, the Poisson distribution correctly models one of the main noise sources of digital images acquisition: the photon shot noise.

3.2 Performance Bounds

This section is devoted to the irradiance estimation problem and to the computation of the corresponding performance bounds under the acquisition model introduced in Chapter 2 (Section 2.1.4). Let us recall the basics of the problem. For each pixel of the sensor, we observe the vector variable $\mathrm{z} = \mathrm{z}_1, \ldots, \mathrm{z}_T$ corresponding to different exposure times $\tau_1, \ldots, \tau_T$. Each z_i is a realization of a random variable Z_i following Model (2.2)

$$\mathrm{Z}_i \sim \mathcal{N}(ga\tau_i C + \mu_R, g^2 a\tau_i C + \sigma_R^2),$$

and we wish to estimate the irradiance C at this pixel, *i.e* the average number of photons reaching the pixel per unit time. This estimation will be done on a per-pixel basis, assuming that images are perfectly registered. Throughout this section, we assume that saturated samples have been discarded, as is done in all existent irradiance estimation methods. The inclusion of saturated samples and its impact on the estimators performances will be studied in Section 3.3.

After an introduction on the irradiance estimation as a statistical problem, the section addresses the computation of the Cramér-Rao lower bound (CRLB) for this model and the question of its attainability. We then proceed with the study of the comparative performance of existent estimators against the CRLB. We focus in particular on the MLE since it was found to outperform experimentally other methods in [Kirk and Andersen 2006; Granados et al. 2010].

3.2.1 Irradiance estimation: a statistical problem

Finding the optimal estimator $\hat{C}$ of C from the observations z_i is not obvious in practice. By "optimal", we mean that the estimator $\hat{C}$ should minimize the quadratic risk (or mean square error) $\mathrm{E}[|\hat{C} - C|^2]$. The main difficulties in this estimation come from the small sample size (between 2 and 6 samples in practice), and from the low SNR of the samples when C is low.

Biased or not biased? The MSE of $\hat{C}$ can be decomposed as the sum of its squared bias and its variance. One question that should be raised is thus the one of the right balance between these two terms. As underlined in the introduction, most of the irradiance estimators proposed in the literature consist of linear combinations of the values $\frac{\mathrm{z}_i}{ga\tau_i}$, with weights w_i summing to 1. In most cases, the weights depend on the input samples, thus the estimators are *a priori* biased. However, since $\mathrm{E}\left[\frac{\mathrm{Z}_i}{ga\tau_i}\right] = C$, they are unbiased to a first order approximation [2]. Only two methods differ from the others in this regard: the approach by Debevec and Malik [1997], which is also unbiased to a first order approximation, but only as an estimator of $\ln C$, and the MLE-based approach by Granados et al. [2010], which is unbiased only asymptotically. In practice, we checked that for all the tested estimators, the ratio $\frac{(\mathrm{E}[|\hat{C}-C|])^2}{C^2}$ remains very close to 0 whatever the value of C (see

[2] The delta method [Rice 1995] can be used to compute a first order approximation of the expected value of the estimators.

Section 3.2.4). This suggests that for all these estimators, the MSE is strongly dominated by the variance. As a consequence, in the following we mainly focus on unbiased (or with negligible biases) estimators.

Is MLE a good estimator of C? The second question coming naturally to mind concerning $\hat{C}$ is the one of the optimality of the MLE. Indeed, the MLE is known to be asymptotically efficient. However, in the HDR imaging problem the sample size is too small (normally in the order of 2 to 6) to consider the asymptotic approximation. Yet, even with a finite number of samples (less than 6 in our case), it can be shown that the MLE would be an optimal solution in terms of MSE if the variance $g^2 aC\tau_i + \sigma_R^2$ in Model (2.2) were a constant or were proportional to C.

The constant variance approximation is valid in very low irradiance conditions ($g^2 aC\tau_i \ll \sigma_R^2$), resulting in a linear MLE, more precisely, it is the weighted average of the irradiance estimations for each exposure. In high irradiance conditions, the variance can be approximated as proportional to C ($g^2 aC\tau_i \gg \sigma_R^2$). The MLE for that case is not linear but has a closed-form. However, under Model (2.2), the MLE does not have a closed-form. The solutions proposed in the literature to compute it numerically consist in simplifying the model [Kirk and Andersen 2006] and making use of an iterative approach [Granados et al. 2010]. Nevertheless, the quality of the estimation in this general case cannot *a priori* be easily stated. One of the goals of this section is to answer this question.

3.2.2 Cramér-Rao lower bound for irradiance estimation

The previous paragraphs motivate the study of the performance bounds of the estimation problem. Once the expected performance limits are known, we may determine whether they can be reached, and compare the results obtained by existing estimation methods against them. This comparison states how close existing methods are to the limits, and allows to quantify how much room is left for improvement.

The performance bound of the problem is given by the Cramér-Rao lower bound theorem. The Cramér-Rao lower bound states the minimum variance that we can expect to achieve (for a given bias). Knowing that the samples distribution is given by (2.2), we compute the Cramér-Rao lower bound for unbiased estimators.

Proposition 1. *Given $\mathrm{Z}_1, \ldots, \mathrm{Z}_T$, independent Gaussian distributed random variables with mean and variance given by $\mu_i = ga\tau_i C$ and $\sigma_i^2 = g^2 a\tau_i C + \sigma_R^2$ (Model (2.2)), the variance of any unbiased estimator $\hat{C}$ of the irradiance parameter C from $\mathrm{z}_1, \ldots, \mathrm{z}_T$ observations of these random variables verifies*

$$var(\hat{C}) \geq \textsc{crlb} = \left[\sum_{i=1}^{T} \frac{(ga\tau_i)^2}{g^2 a\tau_i C + \sigma_R^2} + \frac{(g^2 a\tau_i)^2}{2(g^2 a\tau_i C + \sigma_R^2)^2} \right]^{-1}, \tag{3.2}$$

where $\tau_1, \ldots, \tau_T$ are the considered exposure times.

Proof. The logarithm of the probability density function of $\mathrm{Z} = (\mathrm{Z}_1, \ldots, \mathrm{Z}_T)$ is

$$\ln p(\mathrm{Z}, C) = -\frac{1}{2} \sum_{i=1}^{T} \ln(2\pi) + \ln(g^2 a\tau_i C + \sigma_R^2) + \frac{(\mathrm{Z}_i - ga\tau_i C)^2}{g^2 a\tau_i C + \sigma_R^2}. \tag{3.3}$$

The second derivative is given by

$$\frac{\partial^2 \ln p(\mathrm{Z}, C)}{\partial C^2} = \frac{1}{2}\sum_{i=1}^{T} \frac{(ga\tau_i)^2(g^4 a\tau_i C - 2g^2\mathrm{Z}_i^2 - 4g\sigma_R^2\mathrm{Z}_i + g^2\sigma_R^2 - 2\sigma_R^4)}{(g^2 a\tau_i C + \sigma_R^2)^3}. \tag{3.4}$$

The CRLB is defined as

$$\text{CRLB} = \frac{1}{I(C)} = \frac{-1}{E\left[\frac{\partial^2 \ln p(\mathrm{Z},C)}{\partial C^2}\right]}. \tag{3.5}$$

Using the linearity of the expectation, and that $\mathrm{E}(\mathrm{Z}_i) = ga\tau_i C$, $\mathrm{var}(\mathrm{Z}_i) = g^2 a\tau_i C + \sigma_R^2$, it follows that

$$\mathrm{E}\left[\frac{\partial^2 \ln p(\mathrm{Z}, C)}{\partial C^2}\right] = -\left(\frac{g^2}{2}\sum_{i=1}^{T} \frac{2a^2\tau_i^2}{g^2 a\tau_i C + \sigma_R^2} + \frac{(ga\tau_i)^2}{(g^2 a\tau_i C + \sigma_R^2)^2}\right). \tag{3.6}$$

Then,

$$\text{CRLB} = \left[\sum_{i=1}^{T} \frac{(ga\tau_i)^2}{g^2 a\tau_i C + \sigma_R^2} + \frac{(g^2 a\tau_i)^2}{2(g^2 a\tau_i C + \sigma_R^2)^2}\right]^{-1}. \tag{3.7}$$

□

The MSE of any unbiased estimator is bounded by this expression. Notice that for high irradiance values, where σ_R^2 can be neglected, the CRLB does not depend on the specific exposure times but on the total acquired time, i.e., the sum of the exposure times.

The immediate question to be raised is whether an efficient estimator exists (*i.e.* an unbiased estimator that attains the CRLB). The Cramér-Rao lower bound theorem states that an unbiased estimator may be found that attains the bound for all C if and only if the first derivative of the probability density function can be factorized as

$$\frac{\partial \ln p(\mathrm{Z}, c)}{\partial c} = I(c)(h(\mathrm{Z}) - c), \tag{3.8}$$

for some functions h and I. In that case, the efficient estimator is $\hat{C} = h(\mathrm{Z})$, and its variance is $1/I(C)$ [Kay 1993]. Equation (3.8) must be valid for any c and any Z.

Proposition 2. *For given $\mathrm{z}_1, \ldots, \mathrm{z}_T$, independent observations of the Gaussian distributed random variables with $\mu_i = ga\tau_i C$ and $\sigma_i^2 = g^2 a\tau_i C + \sigma_R^2$ (Model (2.2)) we have*

$$\lim_{c\to\infty} \frac{\partial \ln p(\mathrm{z}, c)}{\partial c} = -\frac{1}{2}\sum_{i=1}^{T} a\tau_i, \tag{3.9}$$

$$\lim_{c\to\infty} I(c)(h(\mathrm{z}) - c) = -\sum_{i=1}^{T} a\tau_i \quad \forall h(\mathrm{z}). \tag{3.10}$$

Proof. The log-likelihood function of z is given by (3.3) and its first derivative is

$$\frac{\partial \ln p(\mathrm{z}, c)}{\partial c} = -\frac{1}{2}\sum_{i=1}^{T} \frac{a\tau_i[(g^2 a\tau_i c + \sigma_R^2)^2 - (g\mathrm{z}_i + \sigma_R^2)^2 + g^2 a\tau_i(g^2 a\tau_i c + \sigma_R^2)]}{(g^2 a\tau_i c + \sigma_R^2)^2}. \quad (3.11)$$

Thus, for the given z,

$$\lim_{c\to\infty} \frac{\partial \ln p(\mathrm{z}, c)}{\partial c} = -\frac{1}{2}\sum_{i=1}^{T} a\tau_i. \quad (3.12)$$

Also,

$$\lim_{c\to\infty} I(c)(h(\mathrm{z}) - c) = \lim_{c\to\infty} \left(\sum_{i=1}^{T} \frac{(ga\tau_i)^2}{g^2 a\tau_i c + \sigma_R^2} + \frac{(g^2 a\tau_i)^2}{2(g^2 a\tau_i c + \sigma_R^2)^2} \right) (h(\mathrm{z}) - c). \quad (3.13)$$

Now, since the function $h(\mathrm{z})$ does not depend on c,

$$\lim_{c\to\infty} I(c)h(\mathrm{z}) = \lim_{c\to\infty} \left(\sum_{i=1}^{T} \frac{(ga\tau_i)^2}{g^2 a\tau_i c + \sigma_R^2} + \frac{(g^2 a\tau_i)^2}{2(g^2 a\tau_i c + \sigma_R^2)^2} \right) h(\mathrm{z}) \quad (3.14)$$

$$= 0 \quad (3.15)$$

and

$$\lim_{c\to\infty} I(c)c = \lim_{c\to\infty} \sum_{i=1}^{T} \frac{(ga\tau_i)^2 c}{g^2 a\tau_i c + \sigma_R^2} + \frac{(g^2 a\tau_i)^2 c}{2(g^2 a\tau_i c + \sigma_R^2)^2} \quad (3.16)$$

$$= \sum_{i=1}^{T} a\tau_i. \quad (3.17)$$

Then

$$\lim_{c\to\infty} I(c)(h(\mathrm{z}) - c) = -\sum_{i=1}^{T} a\tau_i. \quad (3.18)$$

□

The results in (3.9) and (3.10) are different, which proves that an efficient estimator does not exist for the problem.

Once the performance bounds of the problem have been determined, we are interested in the comparative performance of existing irradiance estimation methods against the CRLB. As previously stated, we focus on the MLE since it was found to outperform experimentally other methods in [Kirk and Andersen 2006; Granados et al. 2010].

The MLE for the irradiance under Model (2.2) does not have a closed-form. As a consequence, it is not obvious to directly compute its variance to evaluate its performance against the CRLB. Instead, we propose in Section 3.2.4 a detailed experimental study of its performance, together with that of other widely known irradiance estimation methods, relatively to this bound. Before the experimental study we present in Section 3.2.3

some theoretical hints of the optimality of the MLE. This result is verified experimentally afterward in Section 3.2.4.

It is interesting to remark that the previous bound can be directly extended to the case of variable gain g. In this work we focus on the approach of HDR image generation by the combination of images acquired with different exposure times, while keeping the same gain parameter for all images (fixed ISO setting). However, another possible approach for the generation of HDR images is the fusion of images acquired with different exposure times and gain settings [Hasinoff et al. 2010]. The performance bound in that case is

$$\text{CRLB} = \left[\sum_{i=1}^{T} \frac{(g_i a \tau_i)^2}{g_i^2 a \tau_i C + \sigma_{R_i}^2} + \frac{(g_i^2 a \tau_i)^2}{2(g_i^2 a \tau_i C + \sigma_{R_i}^2)^2} \right]^{-1}, \tag{3.19}$$

where the i-th image is acquired with exposure time τ_i and gain value g_i. The readout noise $\sigma_{R_i}^2$ changes with the ISO setting since part of this noise is amplified by the gain.

Regarding biased estimators, the performance bound for a given bias $b(C)$ can be obtained multiplying (3.2) by $(1 + \frac{\partial b(C)}{\partial C})^2$, where $\frac{\partial b(C)}{\partial C}$ is the first derivative of the bias.

3.2.3 How close is the MLE from the CRLB: Some theoretical hints

The lack of closed-form for the MLE under Model (2.2) motivates Granados et al. [2010] to propose an iterative algorithm to approximate the MLE solution. We show in this section that the variance of the first iteration of this iterative procedure is very close to the CRLB. Let us begin by showing that this iterative procedure actually converges to the MLE solution.

Convergence of Granados et al. iterative procedure Let us consider T observations $\mathrm{z}_1, \ldots, \mathrm{z}_T$ of the independent random variables $\mathrm{Z}_1, \ldots, \mathrm{Z}_T$ following Model (2.2). The corresponding log-likelihood function is given by (3.3) and its first derivative is (3.11). Let C_0 be the irradiance value that vanishes (3.11). From (3.11) we have,

$$\sum_{i=1}^{T} a\tau_i \left[\frac{g^2 a\tau_i C_0 + \sigma_R^2}{g^2 a\tau_i C_0 + \sigma_R^2} - \frac{(g\mathrm{z}_i + \sigma_R^2)^2}{(g^2 a\tau_i C_0 + \sigma_R^2)^2} + \frac{g^2 a\tau_i}{g^2 a\tau_i C_0 + \sigma_R^2} \right] = 0. \tag{3.20}$$

Thus,

$$C_0 \sum_{i=1}^{T} \frac{(ga\tau_i)^2}{g^2 a\tau_i C_0 + \sigma_R^2} = \sum_{i=1}^{T} a\tau_i \left[\frac{(g\mathrm{z}_i + \sigma_R^2)^2}{(g^2 a\tau_i C_0 + \sigma_R^2)^2} - \frac{g^2 a\tau_i + \sigma_R^2}{g^2 a\tau_i C_0 + \sigma_R^2} \right], \tag{3.21}$$

$$C_0 = \frac{\sum_{i=1}^{T} \frac{(ga\tau_i)^2}{g^2 a\tau_i C_0 + \sigma_R^2} \left[\frac{(g\mathrm{z}_i + \sigma_R^2)^2}{(g^2 a\tau_i C_0 + \sigma_R^2) g^2 a\tau_i} - \frac{g^2 a\tau_i + \sigma_R^2}{g^2 a\tau_i} \right]}{\sum_{i=1}^{T} \frac{(ga\tau_i)^2}{g^2 a\tau_i C_0 + \sigma_R^2}}. \tag{3.22}$$

The pixel values z_i can be expressed as

$$\mathrm{z}_i = ga\tau_i C_0 + \varepsilon_i, \tag{3.23}$$

where ε_i is an error term that according to Model (2.2) takes values with an order of magnitude given by $\sqrt{g^2 a\tau_i C_0 + \sigma_R^2}$. Thus, the term in square brackets can be rewritten as,

$$\frac{(g\mathrm{z}_i + \sigma_R^2)^2}{(g^2 a\tau_i C_0 + \sigma_R^2) g^2 a\tau_i} - \frac{g^2 a\tau_i + \sigma_R^2}{g^2 a\tau_i} = \frac{1}{g^2 a\tau_i}\left[\frac{(g\mathrm{z}_i + \sigma_R^2)(g^2 a\tau_i + g\varepsilon_i + \sigma_R^2)}{(g^2 a\tau_i C_0 + \sigma_R^2) g^2 a\tau_i}\right. \tag{3.24}$$

$$\left. - g^2 a\tau_i + \sigma_R^2\right] \tag{3.25}$$

$$= \frac{1}{g^2 a\tau_i}\left[g(\mathrm{z}_i - g a\tau_i) + \frac{g\varepsilon_i(g\mathrm{z}_i + \sigma_R^2)}{g^2 a\tau_i C_0 + \sigma_R^2}\right] \tag{3.26}$$

$$= \mathrm{x}_i - 1 + \frac{\varepsilon_i}{g a\tau_i} + \frac{\varepsilon_i^2}{a\tau_i(g^2 a\tau_i C_0 + \sigma_R^2)} \tag{3.27}$$

$$\approx \mathrm{x}_i. \tag{3.28}$$

where x_i are the pixel values normalized to the irradiance domain according to $\mathrm{x}_i = \frac{\mathrm{z}_i}{g a\tau_i}$. The approximation is valid since, in general, the irradiance values verify $\mathrm{x}_i \gg 1$ and according to Model (2.2) ε_i takes values with an order of magnitude given by $\sqrt{g^2 a\tau_i C_0 + \sigma_R^2}$. Hence, from (3.22) and (3.28) we have

$$C_0 \approx \frac{\sum_{i=1}^{T} \frac{(g a\tau_i)^2}{g^2 a\tau_i C_0 + \sigma_R^2} \mathrm{x}_i}{\sum_{i=1}^{T} \frac{(g a\tau_i)^2}{g^2 a\tau_i C_0 + \sigma_R^2}}, \tag{3.29}$$

which is the fixed point iteration proposed by Granados et al. [2010][3] . This formula is also obtained if we compute the first derivative of (3.3) neglecting the dependence of the pixel's variance on the irradiance C.

Now, let us consider the function $f : \mathbb{R} \to \mathbb{R}$, given by

$$f(C) = \sum_{i=1}^{T} w_i(C)\mathrm{x}_i \quad \text{with } w_i(C) = \frac{\frac{(g a\tau_i)^2}{g^2 a\tau_i C + \sigma_R^2}}{\sum_{h=1}^{T} \frac{(g a\tau_h)^2}{g^2 a\tau_h C + \sigma_R^2}}. \tag{3.30}$$

Proposition 3. *For $C \geq 0$, $f(C)$ is bounded by*

$$\mathrm{x}_m \leq f(C) \leq \mathrm{x}_M, \tag{3.31}$$

with x_m and x_M the minimum and maximum values of $\mathrm{x}_i = \frac{\mathrm{z}_i}{g a\tau_i} \forall i = 1, \ldots, T$.

Proof. For the upper bound, since $C \geq 0$, we have

$$f(C) = \frac{\sum_{i=1}^{T} \frac{(g a\tau_i)^2 \mathrm{x}_i}{g^2 a\tau_i C + \sigma_R^2}}{\sum_{i=1}^{T} \frac{(g a\tau_i)^2}{g^2 a\tau_i C + \sigma_R^2}} \leq \frac{\sum_{i=1}^{T} \frac{(g a\tau_i)^2 \mathrm{x}_M}{g^2 a\tau_i C + \sigma_R^2}}{\sum_{i=1}^{T} \frac{(g a\tau_i)^2}{g^2 a\tau_i C + \sigma_R^2}} = \frac{\mathrm{x}_M \sum_{i=1}^{T} \frac{(g a\tau_i)^2}{g^2 a\tau_i C + \sigma_R^2}}{\sum_{i=1}^{T} \frac{(g a\tau_i)^2}{g^2 a\tau_i C + \sigma_R^2}} = \mathrm{x}_M. \tag{3.32}$$

Idem for the lower bound. □

If $C^{(0)} > 0$, $\mathrm{x}_m \leq C^{(0)} \leq \mathrm{x}_M$ and $C^{(j+1)} = f(C^{(j)})$, with f bounded between x_m and x_M, the sequence $C^{(j)}$ is bounded and it has therefore a convergent subsequence.

In theory, the first derivative of f can verify $|f'| > 1$. However, we observe in practice that $|f'| < 1$ in all our experiments and therefore, in those cases, f has a unique fixed point.

[3] Unlike Granados et al. we neglect the dark currents since their contribution to global noise is minimal for the exposures normally used in HDR image generation methods.

Variance of the MLE approximation At iteration $(j+1)$ the irradiance $\hat{C}^{(j+1)}$ is estimated according to

$$\hat{C}^{(j+1)} = \sum_{i=1}^{T} w_i(\hat{C}^{(j)})\mathrm{x}_i \quad \text{with } \mathrm{x}_i = \frac{\mathrm{z}_i}{ga\tau_i}, \quad w_i = \frac{\frac{(ga\tau_i)^2}{g^2 a\tau_i \hat{C}^{(j)} + \sigma_R^2}}{\sum_{h=1}^{T} \frac{(ga\tau_h)^2}{g^2 a\tau_h \hat{C}^{(j)} + \sigma_R^2}}. \tag{3.33}$$

It can be verified that in practice, if the weights are correctly initialized, the estimator remains almost unchanged after the first iteration. A correct initialization is $\hat{C}^{(0)} = \frac{\mathrm{z}_i}{ga\tau_i}$, i.e. the weights are computed directly from the input samples.

Proposition 4. *A first order approximation of the variance of the estimator after the first iteration can be computed as*

$$var(\hat{C}^{(1)}) = \left[\sum_{i=1}^{T} \frac{(ga\tau_i)^2}{g^2 a\tau_i C + \sigma_R^2}\right]^{-1} + E(o((\mathrm{Z}-\mu)^2)). \tag{3.34}$$

Proof. Let $\mathrm{Y} = (\mathrm{Y}_1, \ldots, \mathrm{Y}_T)$ be a random vector, with $\mathrm{Y}_i = \mathrm{Z}_i$. Thus $\mathrm{Y} \sim N(\mu, \mathbf{\Sigma})$ with

$$\mu = (\mu_1, \ldots, \mu_T), \quad \mu_i = g\tau_i aC \quad \forall\, i = 1, \ldots, T, \tag{3.35}$$

$$\mathbf{\Sigma} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_T^2), \quad \sigma_i^2 = g^2\tau_i aC + \sigma_R^2 \quad \forall\, i = 1, \ldots, T. \tag{3.36}$$

Consider the function

$$h(\mathrm{Y}) = \frac{\sum_{i=1}^{T} \frac{g\tau_i a \mathrm{Y}_i}{g\mathrm{Y}_i + \sigma_R^2}}{\sum_{i=1}^{T} \frac{(g\tau_i a)^2}{g\mathrm{Y}_i + \sigma_R^2}}. \tag{3.37}$$

Being Y Gaussian distributed, and noting that h is continuously differentiable and that

$$\sum_{j=1}^{T}\sum_{k=1}^{T} \sigma_{jk} \frac{\partial h(\mu)}{\partial \mathrm{Y}_j} \frac{\partial h(\mu)}{\partial \mathrm{Y}_k} > 0, \tag{3.38}$$

it follows from the Delta Method theorem [Rice 1995] that the variance of $h(\mathrm{Y})$ can be computed to a first order approximation as

$$\mathrm{var}(h(\mathrm{Y})) = \sum_{i=1}^{T} \left[\frac{\partial h(\mathrm{Y})}{\partial \mathrm{Y}_i}\Big|_{\mathrm{Y}=\mu_\mathrm{Y}}\right]^2 \sigma_{\mathrm{Y}_i}^2 + \mathrm{E}(o((\mathrm{Y}-\mu)^2)) \tag{3.39}$$

$$= \sum_{i=1}^{T} \left[\frac{g\tau_i a}{g^2\tau_i aC + \sigma_R^2}\right]^2 \left[\frac{(g\tau_i a)^2}{g^2\tau_i aC + \sigma_R^2}\right]^{-2} (g^2\tau_i aC + \sigma_R^2) + \mathrm{E}(o((\mathrm{Z}-\mu)^2)) \tag{3.40}$$

$$\approx \left[\frac{(g\tau_i a)^2}{g^2\tau_i aC + \sigma_R^2}\right]^{-1}. \tag{3.41}$$

□

Figure 3.1: HDR image taken as ground-truth for the synthetic tests. Dynamic range 12.7 stops. From [Hasinoff et al.].

On the other hand, recall the CRLB formula introduced in (3.2)

$$\text{CRLB} = \left[\sum_{i=1}^{T} \frac{(ga\tau_i)^2}{g^2 a\tau_i C + \sigma_R^2} + \frac{(g^2 a\tau_i)^2}{2(g^2 a\tau_i C + \sigma_R^2)^2} \right]^{-1}. \tag{3.42}$$

Hence the variance of $\hat{C}^{(1)}$ approaches the CRLB if

$$\frac{g^2 a^2 \tau_i^2}{g^2 a\tau_i C + \sigma_R^2} \gg \frac{g^4 a^2 \tau_i^2}{2(g^2 a\tau_i C + \sigma_R^2)^2} \quad \forall i = 1, \ldots, T, \tag{3.43}$$

which can be rewritten as

$$a\tau_i C \gg \frac{1}{2} - \frac{\sigma_R^2}{g^2} \quad \forall i = 1, \ldots, T. \tag{3.44}$$

This inequality is verified in practice, even for small $a\tau_i C$ values, because the term σ_R^2/g^2 is much larger than $1/2$ for most cameras (e.g. for the Canon 400D camera the σ_R^2/g^2 ratio takes values 103 (ISO 200), 55 (ISO 400), 35 (ISO 800)). This suggests that the iterative algorithm proposed by Granados et al. almost achieves the CRLB after the first iteration.

3.2.4 How close are existing estimators from the CRLB: An experimental study

In Section 3.2.2 we showed that an efficient estimator for the irradiance does not exist under Model (2.2). Yet it is of interest to study how close to the CRLB existing methods perform.

In the present section an experimental study is conducted to analyze the performance of various methods widely known for the irradiance map estimation: MLE (Granados et al. [2010]), Kirk and Andersen [2006], Robertson et al. [2003], Debevec and Malik [1997], Mitsunaga and Nayar [1999], Reinhard et al. [2005] and a quite simple weighting scheme, classically used in the imaging industry, which we refer hereafter to as Poisson approach. The estimator by Kirk and Andersen [2006] is computed according to expression (3.1), with the weights given by $w_i^p = \frac{\tau_i^2}{g z_i^p + \sigma_R^2}$. Given that it does not take into account the PRNU factors, the estimator thus obtained is biased. A trivial modification of these

weights allows us to have a non-biased version of the estimator, with $w_i^p = \frac{a^2\tau_i^2}{gz_i^p+\sigma_R^2}$, which we will refer hereafter to as the *Modified Kirk and Andersen's method*. With this modification, we think the comparison with Kirk and Andersen's method becomes more fair. Thus, in the following, we will evaluate the modified version of Kirk and Andersen's method.

The experimental study is performed with synthetic data only since the knowledge of the exact ground-truth is imperative for the computation of the CRLB.

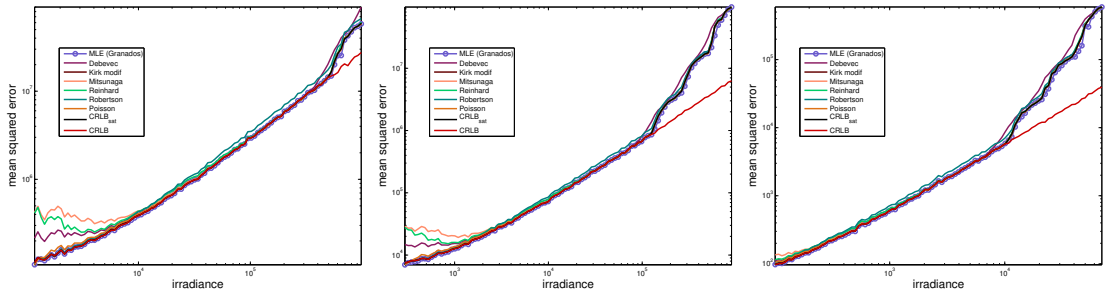
Synthetic data generation Synthetic samples are generated from an HDR image taken as ground-truth assuming the pixel values follow Model (2.2). Figure 3.1 shows a tone mapped version of the ground-truth. Notice that a HDR image is taken as ground-truth in order to consider a realistic dynamic range, and to easily visualize the results. However, given that the estimation is done on a per-pixel basis, it only depends on each individual pixel irradiance, independently of the pixels' ordering or location in the image. In order to evaluate the dependence on the shutter speeds, 6 sets of exposure times are tested:

- Two sets of relatively long exposure times with 4 and 6 elements:
  - $\tau_{4L} = (1, 1/2, 1/4, 1/8)s$,
  - $\tau_{6L} = (1, 1/2, 1/4, 1/8, 1/16, 1/32)s$;
- Two sets of relatively short exposure times with 4 and 6 elements:
  - $\tau_{4S} = (1/50, 1/100, 1/200, 1/400)s$,
  - $\tau_{6S} = (1/50, 1/100, 1/200, 1/400, 1/600, 1/800)s$;
- Two sets of medium exposure times of 4 and 6 elements:
  - $\tau_{4M} = (1/12.4, 1/25, 1/50, 1/100)s$,
  - $\tau_{6M} = (1/6.2, 1/12.4, 1/25, 1/50, 1/100, 1/200)s$.

Two cameras are simulated[4]:

- Camera A, a Canon 7D set to ISO 200 ($g = 0.87; \sigma_R^2 = 31.6; \mu_R = 2046; z_{sat} = 14042$),
- Camera B, a Canon 400D set to ISO 400 ($g = 0.33; \sigma_R^2 = 6.2; \mu_R = 256; z_{sat} = 4056$).

The PRNU factors at each pixel are simulated following a Gaussian distribution with mean 1 and standard deviation 0.01 [Granados et al. 2010]. The dynamic range of the scene is 12.7 stops. The MLE estimation is computed using the implementation provided by Granados et al. [2010]. The estimations are repeated 1000 times for each irradiance level. The MSE and the variance of the estimators are computed from these 1000 repetitions.



(a) Short exposure times. (b) Medium exposure times. (c) Long exposure times.

Figure 3.2: Comparison of the MSE curves and the CRLB for Camera A with 4 exposure times (results are similar for the other configurations). In all cases the MSE of the MLE is very close to the CRLB.

Camera A														
	Granados		Kirk		Robertson		Poisson		Devebec		Mitsunaga		Reinhard	
	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std		
τ_{4S}	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.002	0.004	0.011	0.009	0.025	0.007	0.021
τ_{6S}	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.002	0.020	0.059	0.061	0.181	0.036	0.125
τ_{4M}	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.002	0.003	0.011	0.008	0.024	0.006	0.020
τ_{6M}	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.002	0.019	0.069	0.084	0.311	0.044	0.204
τ_{4L}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
τ_{6L}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.000	0.000
Camera B														
	Granados		Kirk		Robertson		Poisson		Devebec		Mitsunaga		Reinhard	
	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std		
τ_{4S}	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.001	0.003	0.009	0.007	0.019	0.005	0.016
τ_{6S}	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.017	0.054	0.055	0.174	0.031	0.120
τ_{4M}	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.003	0.010	0.007	0.022	0.005	0.018
τ_{6M}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.011	0.038	0.053	0.192	0.024	0.103
τ_{4L}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000
τ_{6L}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.003	0.000	0.001

Table 3.1: Average and standard deviation of the ratio bias^2/C^2. The estimators' bias is negligible with respect to the corresponding irradiance values.

Evaluation Procedure First, we verify the validity of the hypothesis made in Section 3.1, stating that the bias is negligible for all the evaluated methods. Table 3.1 shows the expected ratio of the squared bias and the squared irradiance for all the tested configurations. In all cases, the ratio is very close to zero implying that the bias is negligible with respect to the irradiance values. Hence, the MSE of these methods can be fairly compared against the Cramér-Rao lower bound for unbiased estimators.

The ratios obtained for Debevec and Malik, Mitsunaga and Nayar and Reinhard et al. are slightly above zero for the shorter exposure times sets. These estimators assume that both, the pixel values and the inverse of the camera response function, take positive values only. However, this is not the case for some digital cameras, and in particular for the raw data following Model (2.2). Under this model, the inverse of the camera response function may take negative values after subtracting the mean μ_R for very low irradiance values. In consequence, the less realistic model proposed by these methods introduces some bias in very low light conditions.

Table 3.2 shows the average and standard deviation of the ratio between the MSE of each estimator and the CRLB for all the tested configurations. The CRLB is computed for each pixel according to (3.2). In the CRLB computation we only take into account the exposure times producing non saturated samples, since in practice the samples corresponding to the other exposures would be saturated and therefore discarded (all methods discard the saturated samples). In the following, the CRLB computed from the non-saturated exposures only will be referred to as CRLB_{SAT}, while CRLB refers to the bound computed considering all the exposures (ideal case without saturation).

Results

Maximum Likelihood Estimation For the MLE (Granados et al. implementation, see Equation (3.33)), in all cases the average ratio is very close to 1, with a standard deviation in the order of 5%, meaning that the MSE of the MLE is very close to the CRLB_{SAT}. Figure 3.2 shows the comparison of the MSE curves against the CRLB and the CRLB_{SAT} as a function of the irradiance C for some of the tested configurations (similar results are found for all configurations). In agreement with the results presented in Table 3.2, the MSE of the MLE is very close to the CRLB_{SAT} (almost indistinguishable in these figures).

From Figure 3.2 it can be verified that, as expected, the MSE results for the low irradiance regions are smaller for the long exposure times set than for the short exposures set. What is not a priori expected is that the same behavior is found for the high irradiance range. For high irradiance, the probability of saturation increases for the longer exposures and this may increase the MSE. The reason for the MSE reduction in high irradiance is that, even if the estimation is performed with less samples due to saturation, the shortest exposure of the long exposures set (1/6.2 = 0.16) is roughly 4 times the sum of all the exposures of the short times set (1/50 + 1/100 + 1/200 + 1/400 = 0.0375). As stated in Section 3.2.2, for high irradiance values where σ_R^2 can be neglected, the CRLB does not depend on the specific exposure times but on the total acquired time, i.e., the sum of the exposure

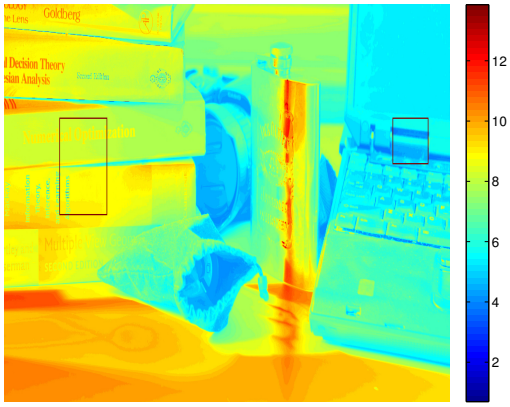
[4]The camera parameters were obtained using the calibration procedure by Granados et al. [2010].

Camera A														
	Granados		Kirk		Robertson		Poisson		Devebec		Mitsunaga		Reinhard	
	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std
τ_{4S}	**0.996**	0.056	1.009	0.049	1.124	0.087	1.034	0.067	1.178	0.280	1.416	0.764	1.307	0.591
τ_{6S}	**0.995**	0.065	1.013	0.056	1.157	0.114	1.051	0.079	1.609	1.059	3.013	3.960	2.164	2.907
τ_{4M}	**0.991**	0.081	1.008	0.050	1.120	0.089	1.028	0.066	1.193	0.304	1.347	0.714	1.257	0.543
τ_{6M}	**0.988**	0.090	1.011	0.055	1.153	0.104	1.047	0.080	1.655	1.251	4.141	7.742	2.660	5.592
τ_{4L}	**0.985**	0.102	1.007	0.049	1.141	0.079	1.011	0.054	1.126	0.296	1.039	0.100	1.056	0.091
τ_{6L}	**0.980**	0.111	1.007	0.051	1.170	0.094	1.025	0.068	1.192	0.297	1.206	0.470	1.071	0.115
Camera B														
	Granados		Kirk		Robertson		Poisson		Devebec		Mitsunaga		Reinhard	
	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std
τ_{4S}	**1.000**	0.051	1.005	0.051	1.116	0.091	1.033	0.071	1.191	0.278	1.380	0.671	1.254	0.489
τ_{6S}	**0.997**	0.053	1.005	0.054	1.144	0.113	1.045	0.080	1.600	1.004	3.013	3.894	2.086	2.823
τ_{4M}	**0.998**	0.051	1.003	0.051	1.111	0.092	1.025	0.069	1.210	0.322	1.328	0.663	1.227	0.497
τ_{6M}	**1.000**	0.055	1.007	0.056	1.145	0.103	1.047	0.083	1.569	0.946	3.651	6.508	2.205	4.104
τ_{4L}	**1.004**	0.046	1.009	0.047	1.131	0.081	1.019	0.055	1.124	0.259	1.058	0.127	1.052	0.086
τ_{6L}	**1.001**	0.050	1.006	0.052	1.160	0.093	1.031	0.073	1.207	0.286	1.281	0.656	1.070	0.130

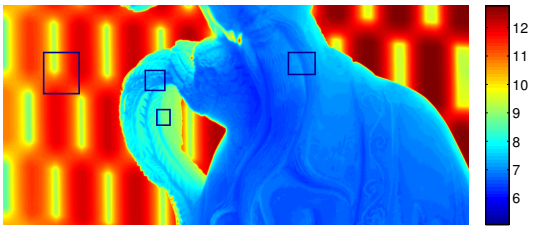
Table 3.2: Average and standard deviation of the ratio $MSE/\mathrm{CRLB}_{\mathrm{SAT}}$ for all tested configurations. For the MLE the ratio is very close to 1, meaning that the MSE of the MLE is very close to the $\mathrm{CRLB}_{\mathrm{SAT}}$. Note that average values for the ratio have an uncertainty given by the reported standard deviation, which justifies average ratio values below 1.

Camera A														
	Granados		Kirk		Robertson		Poisson		Devebec		Mitsunaga		Reinhard	
	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std		
τ_{4S}	0.997	0.057	1.008	0.045	1.170	0.062	1.008	0.046	1.063	0.175	1.020	0.054	1.053	0.074
τ_{6S}	0.997	0.064	1.013	0.047	1.222	0.076	1.022	0.055	1.077	0.140	1.076	0.128	1.075	0.076
τ_{4M}	0.992	0.085	1.010	0.048	1.142	0.078	1.016	0.053	1.111	0.242	1.058	0.127	1.063	0.091
τ_{6M}	0.987	0.096	1.009	0.047	1.193	0.078	1.022	0.058	1.159	0.232	1.125	0.285	1.057	0.078
τ_{4L}	0.974	0.123	0.999	0.048	1.152	0.087	0.997	0.049	1.175	0.361	0.999	0.049	1.023	0.091
τ_{6L}	0.971	0.128	1.001	0.048	1.186	0.098	1.000	0.048	1.187	0.335	1.004	0.050	1.032	0.105
Camera B														
	Granados		Kirk		Robertson		Poisson		Devebec		Mitsunaga		Reinhard	
	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std		
τ_{4S}	1.003	0.047	1.009	0.047	1.160	0.067	1.015	0.050	1.097	0.227	1.030	0.063	1.046	0.068
τ_{6S}	1.001	0.045	1.009	0.045	1.205	0.077	1.030	0.060	1.104	0.178	1.107	0.183	1.064	0.073
τ_{4M}	1.001	0.047	1.007	0.048	1.131	0.082	1.021	0.057	1.141	0.281	1.081	0.175	1.063	0.100
τ_{6M}	1.002	0.050	1.007	0.051	1.180	0.083	1.029	0.070	1.204	0.286	1.159	0.339	1.056	0.095
τ_{4L}	1.002	0.047	1.004	0.048	1.145	0.087	1.003	0.048	1.161	0.315	1.004	0.048	1.019	0.080
τ_{6L}	0.997	0.047	1.000	0.048	1.179	0.096	1.000	0.048	1.187	0.310	1.004	0.050	1.022	0.086

Table 3.3: Average and standard deviation of the ratio $MSE/\mathrm{CRLB}_{\mathrm{SAT}}$ for high irradiance only. The results for all methods are quite close to the $\mathrm{CRLB}_{\mathrm{SAT}}$. Note that average values for the ratio have an uncertainty given by the reported standard deviation, which justifies average ratio values below 1.



(a) Synthetic tests. Image from [Hasinoff et al.].



(b) Real data tests. Images from [Granados et al. 2010].

Figure 3.3: Logarithm of the ground-truth images with selected areas for the synthetic and real data tests.

times.

Modified Kirk and Andersen The Modified Kirk and Andersen estimator is computed according to expression (3.1), with the weights given by $w_i^p = \frac{a^2\tau_i^2}{gz_i^p + \sigma_R^2}$. Recall that for all methods, we assume a linear camera response function f. The results presented on Table 3.2 show that this estimator, like the MLE, performs very close to the CRLB_{SAT}. This is expected since this modified version of the Kirk and Andersen's estimator equals the first iteration of the estimator proposed by Granados et al., given by (3.33).

Robertson et al. The results for the method by Robertson et al. are in the order of 15% above the CRLB_{SAT}. Its performance decreases for the longer exposure time sets. The irradiance is computed according to Equation (3.1) with the weights set to $w_i^p = \tau_i^2$ (and again a linear camera response function). Hence, for the longer exposure time sets it gives low importance to exposures long enough to better contribute to the estimation. It is interesting to recall that this estimator is the MLE when assuming a noise model of constant variance.

Poisson The estimator here named Poisson is the minimum variance unbiased estimator, which also matches the MLE, when the input samples are assumed to follow a Poisson distribution with mean and variance depending on the irradiance parameter. Hence, this is the optimal estimator when neglecting all the other noise sources except the shot noise. This estimator is obtained by setting $w_i^p = \tau_i$. For high irradiance values, where the main noise source is the shot noise, the results obtained are quite close to those obtained by the MLE and thus quite close to the CRLB_{SAT}. However, its performance is degraded in low irradiance where the readout noise is not negligible.

Debevec and Malik, Mitsunaga and Nayar, Reinhard et al. The results obtained for Debevec and Malik, Mitsunaga and Nayar and Reinhard et al. are considerably far from the CRLB_{SAT}. This is mainly due to their poor performance at low irradiance, as shown in Figure 3.2. Indeed, these weighting schemes are highly sensitive to noise in the input samples (e.g. for the linear camera case, Mitsunaga and Nayar's weights are the pixel values) and therefore their performance is severely degraded for low irradiance where the SNR of the samples is lower.

As can be verified in Figure 3.2, their performance highly improves for the high irradiance range. Thus it is interesting to make a more local analysis of performance and compute the mean ratio without considering the low irradiance range. These results are presented in Table 3.3. The results are considerably better in high irradiance.

Debevec and Malik use a hat function to weigh the samples from Equation (3.1):

$$w_i^p = \begin{cases} \mathrm{z}_i^p - \mathrm{z}_{\min} & \text{if } \mathrm{z}_i^p \leq \mathrm{z}_{\text{med}} \\ \mathrm{z}_{\max} - \mathrm{z}_i^p & \text{if } \mathrm{z}_i^p > \mathrm{z}_{\text{med}}, \end{cases} \tag{3.45}$$

with $\mathrm{z}_{\min}$ and $\mathrm{z}_{\max}$ the minimum and maximum pixel values respectively and $\mathrm{z}_{\text{med}} = (\mathrm{z}_{\max} + \mathrm{z}_{\min})/2$. For short exposures its weighting scheme is linearly increasing with the exposure. The decreasing part of the hat function is applied for longer exposures. Thus its performance decreases for the longer exposure times sets since a smaller weight is assigned to longer exposures which yet have the higher SNR. The opposite behavior is found for Mitsunaga and Nayar and Reinhard et al., since their weighting schemes are SNR based. For Mitsunaga and Nayar weights are given by $w_i^p = \mathrm{z}_i^p$, while Reinhard et al. use $w_i^p = \mathrm{z}_i^p(1 - (\mathrm{z}_i^p/\mathrm{z}_{\text{med}} - 1)^{12})$. Their performance increases for longer exposure time sets.

Summary of results

The experimental analysis carried out in this section shows that the MLE performs quite close to the CRLB_{SAT} for all irradiance values and for small sample sized datasets. It is important to remark that these results are for the small sample size case, since we have at most (none of the samples saturated) 4 or 6 samples for the estimation on each pixel. The performance of the MLE close to the CRLB_{SAT} was predicted in Section 3.2.3, where it was shown that the variance of the irradiance estimator almost attained the bound after the first iteration.

Moreover, even if the $\hat{C}_{\text{MLE}}$ is not unbiased, its bias is negligible with respect to the irradiance and its performance is very close to the best we can do among unbiased estimators. It would be interesting to evaluate the performance of other biased estimators. However, there are no *a priori* hints on the parametric form that such a bias could take. This analysis also confirms the fact that the MLE outperforms other estimation methods, as observed in [Kirk and Andersen 2006; Granados et al. 2010].

Finally, it is important to remark that as expected, the CRLB using all samples is below the CRLB_{SAT} (see Figure 3.2). This shows that discarding the saturated samples has a high impact on the bounds of estimation performance.

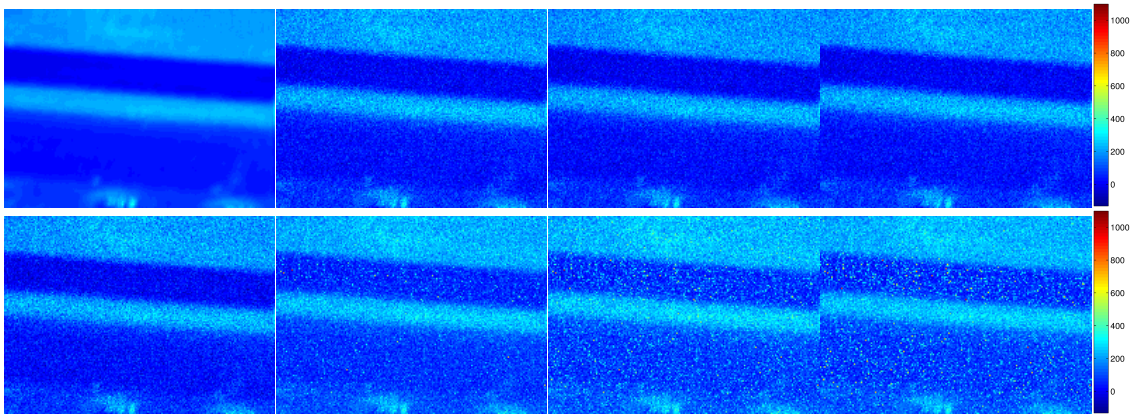


Figure 3.4: **Synthetic data. Estimation in low-level irradiance range.** Patch size 141×193. **Top row:** Ground-truth, MLE, Modified Kirk and Andersen, Robertson et al. **Bottom row:** Poisson, Debevec and Malik, Mitsunaga and Nayar, Reinhard et al. The results for MLE, Modified Kirk and Andersen, Robertson et al. and Poisson are quite similar and less noisy than those of Debevec and Malik, Mitsunaga and Nayar and Reinhard et al.

3.2.5 Visual quality of the estimators

In Section 3.2.4 we verified two facts: first the MLE performs quite close to the CRLB and second it outperforms other irradiance estimation methods. In this section we compare the visual quality of the results obtained by these methods. Is the performance difference among them noticeable in practice? To answer this question we compare their estimations in sub-regions of a test image using both synthetic and real data.

Synthetic Data Figure 3.3a shows a sub-region of the logarithm of the ground-truth image used for the synthetic experiments. The ground-truth image and the presented results correspond to the green channel of the image in Figure 3.1, yet the results are also valid for the red and blue channels. The results obtained for Camera A and τ_{4M} for the two marked sub-regions are displayed in Figures 3.4 and 3.5.

The results presented in Figure 3.4 correspond to a region of low irradiance. In agreement with the results presented in Section 3.2.4, the estimates produced by MLE, Modified Kirk and Andersen, Robertson et al. and Poisson are quite close and less noisy than those obtained by Debevec and Malik, Mitsunaga and Nayar and Reinhard et al. (see Figure 3.2b). Figure 3.5 presents the results observed in high irradiance. The difference between all estimators is hardly noticeable.

Real Data Tests using real data are carried out using the images provided by Granados et al.. The set of images was acquired with a Canon PowerShot S5 camera, using exposures $\tau = (1/1.5, 1/6, 1/25, 1/100, 1/400, 1/1600)s$. The camera configuration is provided in [Granados et al. 2010]. The ground-truth image, shown in Figure 3.3b, is generated using Granados et al. estimator on a set of almost noise-free images, obtained as the average of 36 frames. For more details see [Granados et al. 2010]. The presented results correspond to the red channel, but similar results and considerations hold for the green and blue channels.

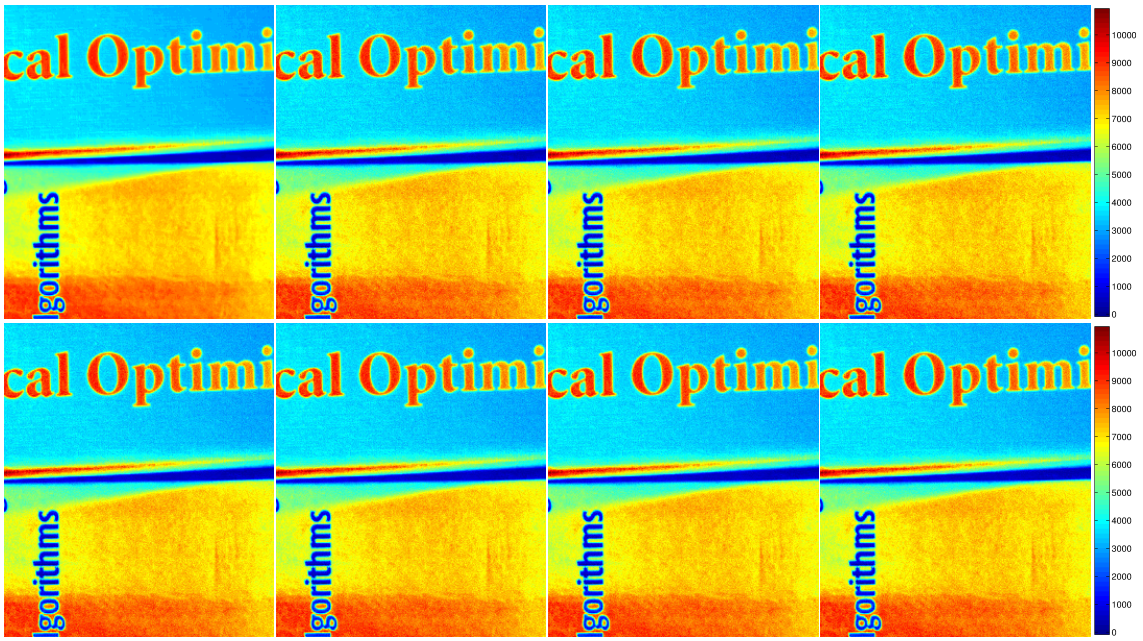


Figure 3.5: **Synthetic data. Estimation in high irradiance range.** Patch size 301×259. **Top row:** Ground-truth, MLE, Modified Kirk and Andersen, Robertson et al. **Bottom row:** Poisson, Debevec and Malik, Mitsunaga and Nayar, Reinhard et al. The difference between all estimators is hardly noticeable.

Figure 3.6 shows examples for irradiance values in the low - mid-level range ranges. As expected, the results for MLE, Modified Kirk and Andersen, Robertson et al. and Poisson are quite similar while the results for Debevec and Malik and Reinhard et al. are better than those of Mitsunaga and Nayar (see Figure 3.2b).

The examples in Figures 3.7 and 3.8 correspond to regions of mid-level and high irradiance. In both cases, as expected, all methods perform quite similarly (see Figure 3.2b).

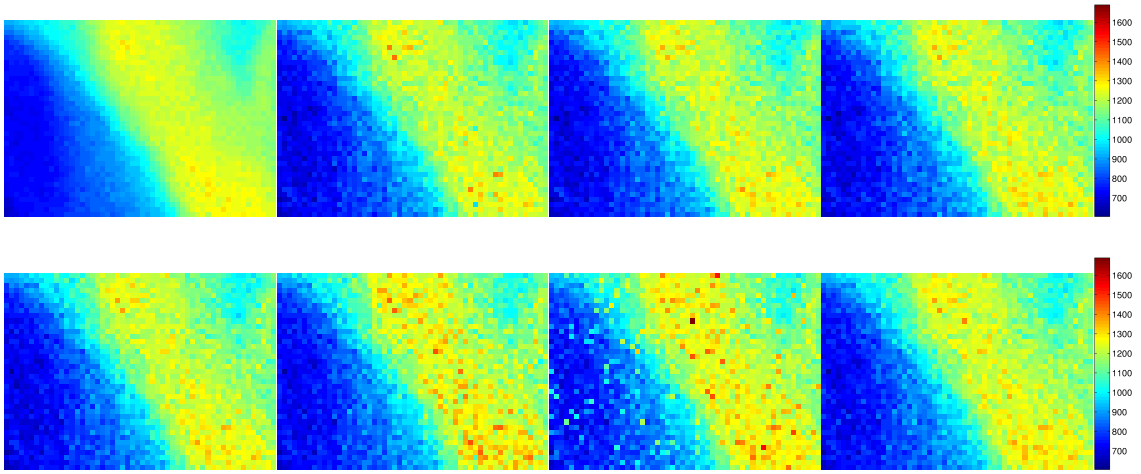
3.3 Including saturation information

To the best of our knowledge, all approaches to irradiance estimation discard saturated samples. However, the experiments presented in Section 3.2.4 show a great lost in performance when comparing the CRLB obtained from only non-saturated samples, with the one obtained from all the samples. We therefore propose a method to include information provided by the saturated samples in the irradiance estimation process.

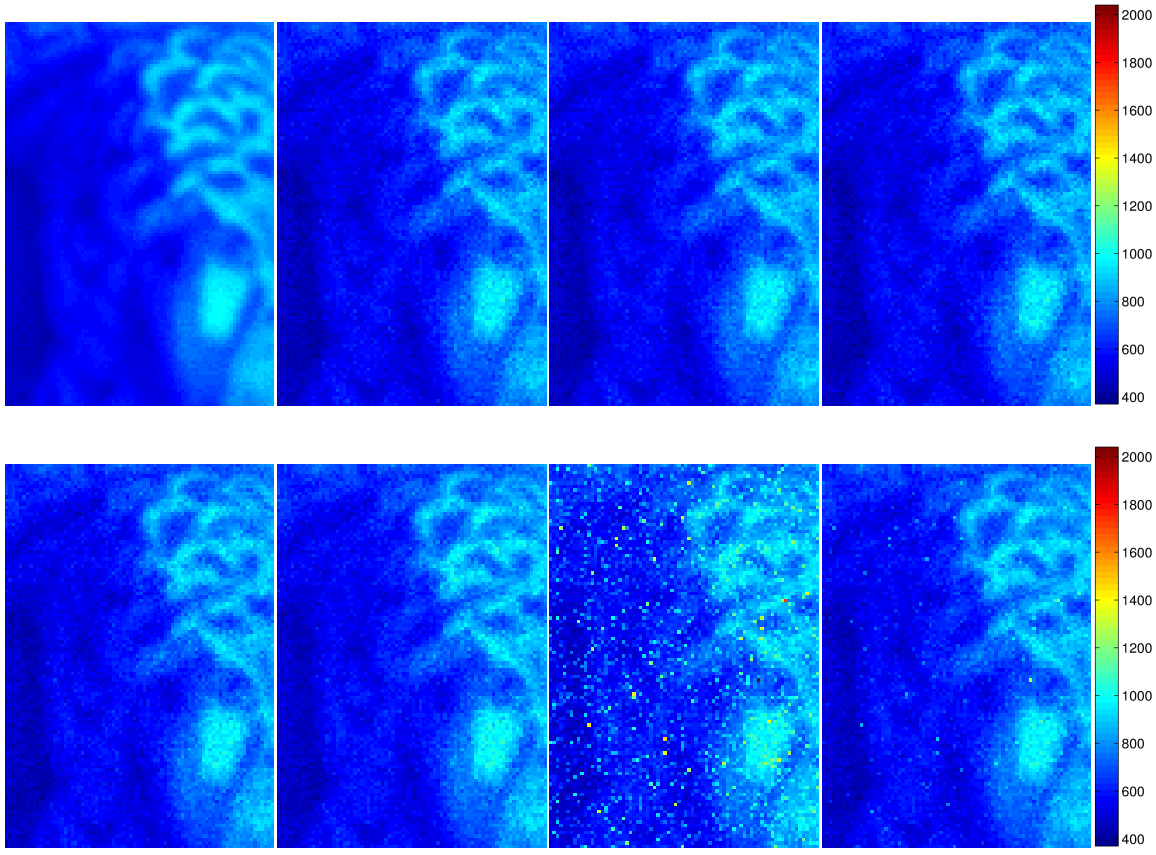
3.3.1 Methodology

Although we do not know the exact value leading to a saturated sample, there is some useful information given by the fact that this value exceeds a given threshold z_{sat}. In this section, we develop a variation of the classical MLE that includes extra information provided by the saturated samples.

Let $Z_1, \ldots, Z_T$ be T independent but not identically distributed random variables, corresponding to the raw values observed at a given pixel p for the different exposure times



(a) Patch size 39×54. **Top row:** Ground-truth, MLE, Modified Kirk and Andersen, Robertson et al. **Bottom row:** Poisson, Debevec and Malik, Mitsunaga and Nayar, Reinhard et al.



(b) Patch size 111×78. **Top row:** Ground-truth, MLE, Modified Kirk and Andersen, Robertson et al. **Bottom row:** Poisson, Debevec and Malik, Mitsunaga and Nayar, Reinhard et al.

Figure 3.6: **Real data. Estimation in low and mid-level irradiance range.** The results for MLE, Modified Kirk and Andersen and Robertson et al. are quite similar while the results for Debevec and Malik and Reinhard et al. are better than those of Mitsunaga and Nayar.

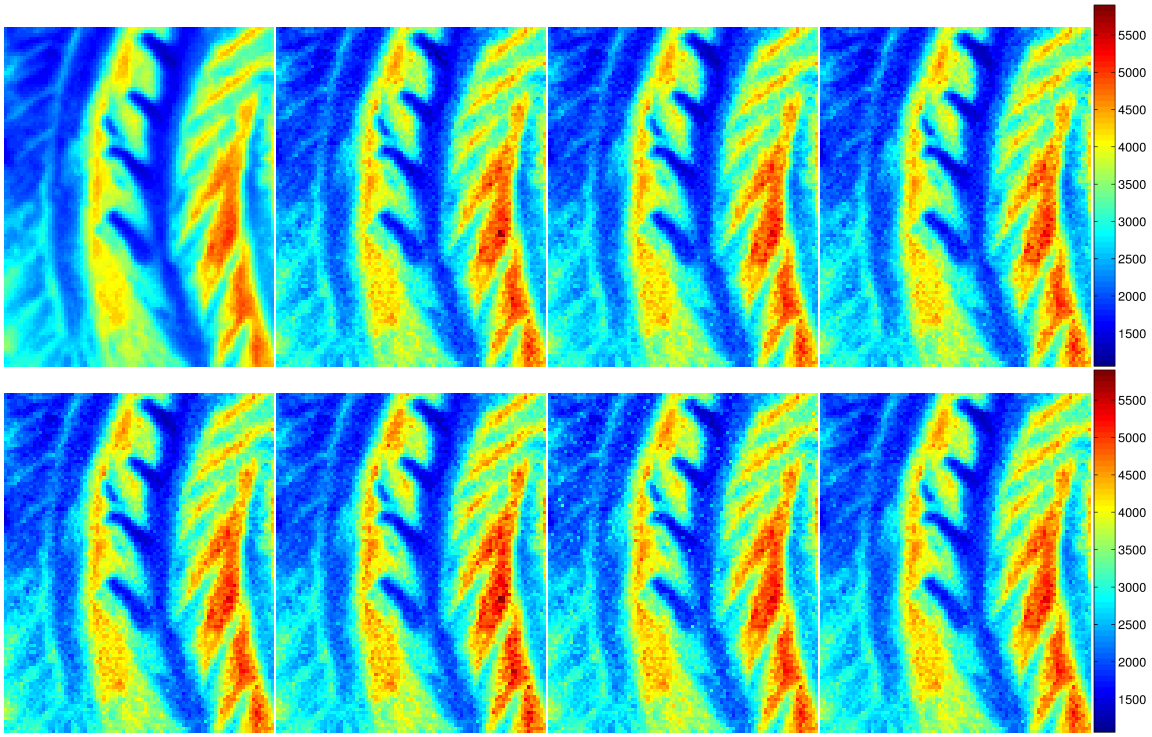


Figure 3.7: **Real data. Estimation in mid-level irradiance range.** Patch size 107×85. **Top row:** Ground-truth, MLE, Modified Kirk and Andersen, Robertson et al. **Bottom row:** Poisson, Debevec and Malik, Mitsunaga and Nayar, Reinhard et al. As expected, results are hardly distinguishable.

$\tau_1, \ldots, \tau_T$. Because of saturation, each Z_j can be written as $\mathrm{Z}_j = \min(\mathrm{X}_j, z_{sat})$, where X_j is a random variable following the law $\mathcal{N}(\mu_j(C), \sigma_j^2(C))$, with $\mu_j(C) = ga\tau_j C$ and $\sigma_j^2(C) = g^2 a\tau_j C + \sigma_R^2$. In the previous sections, saturated samples were discarded and an estimator of the irradiance C was obtained by maximizing the likelihood $\prod_{\mathrm{z}_j < z_{sat}} p_j(\mathrm{z}_j; \phi)$ as a function of ϕ, where $p_j(\mathrm{z}_j; \phi)$ is given by the Gaussian law $\mathcal{N}(\mu_j(\phi), \sigma_j^2(\phi))$. In order to take into account saturated samples, we propose to use instead the full likelihood of the complete set of observations $\mathrm{z} = (\mathrm{z}_1, \ldots, \mathrm{z}_T)$, which writes

$$g(\mathrm{z}|\phi) = \prod_{j=1}^{T} p_j(\mathrm{z}_j; \phi)^{k_j} P_j(\phi)^{1-k_j}, \tag{3.46}$$

where $k_j = 1$ for non-saturated samples, and 0 otherwise. $P_j(\phi)$ is the probability of the j-th sample being saturated, given by

$$P_j(\phi) = \int_{z_{sat}}^{\infty} p_j(\mathrm{z}; \phi) dz. \tag{3.47}$$

Observe that if none of the pixels saturate, $g(\mathrm{z}|\phi)$ (the function defined by Equation (3.46)) is exactly the likelihood maximized by the MLE approach presented in the previous sections.

An efficient way to maximize the likelihood $g(\mathrm{z}|\phi)$ is to use the EM algorithm [Dempster et al. 1977]. In this setting, saturated samples are seen as censored data. Following Dempster et al. [1977], we denote by $\mathrm{x} = (\mathrm{x}_1, \ldots, \mathrm{x}_T)$ the data that would have been

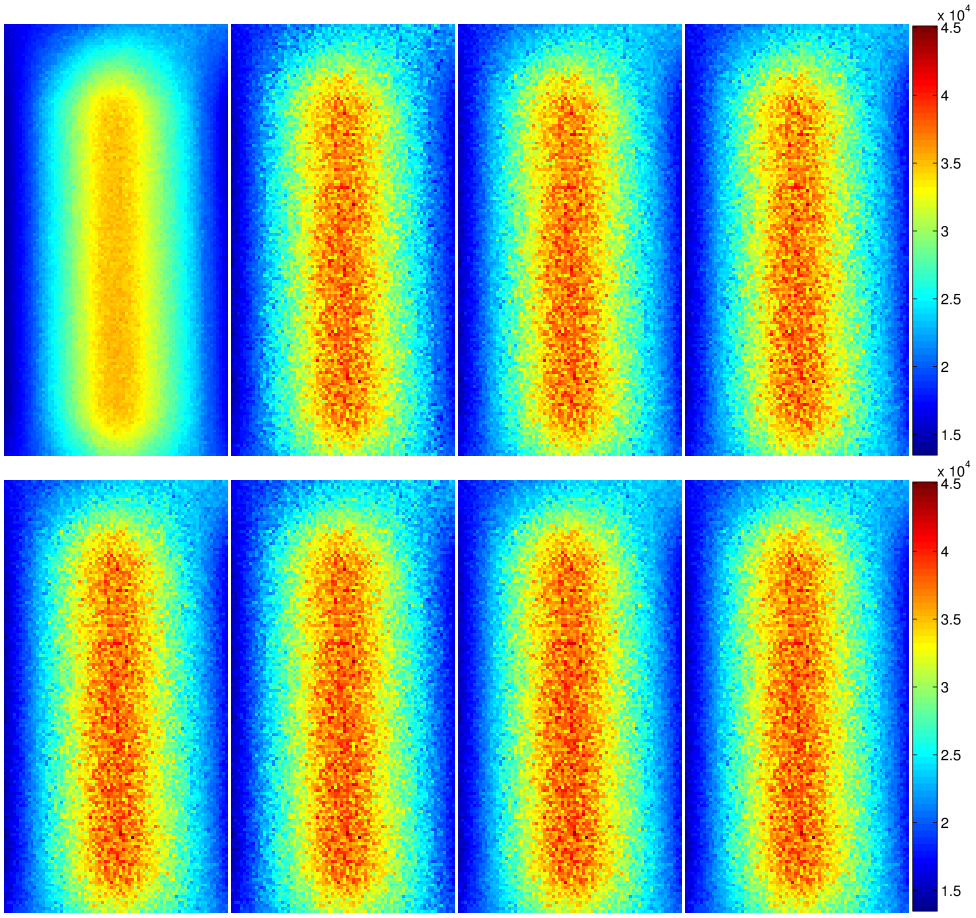


Figure 3.8: **Real data. Estimation in high irradiance range.** Patch size 147×76. **Top row:** Ground-truth, MLE, Modified Kirk and Andersen, Robertson et al. **Bottom row:** Poisson, Debevec and Malik, Mitsunaga and Nayar, Reinhard et al. As expected, results are hardly distinguishable.

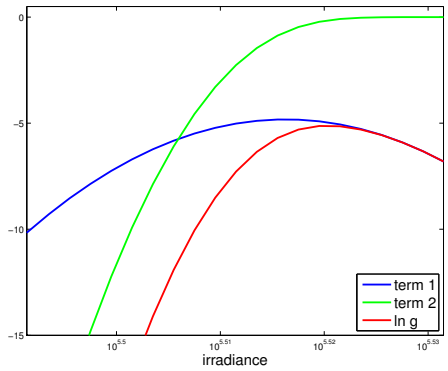


Figure 3.9: In blue the classical log-likelihood curve; in green function (3.63); in red the addition of the two curves: function $\ln g(\mathrm{Z}|\phi)$.

observed if the camera could record beyond the saturation threshold. Each x_j is a realization of the random variable X_j defined above. We define the complete data likelihood of the problem as

$$h(\mathrm{x}|\phi) = \prod_{j=1}^{T} p_j(\mathrm{x}_j;\phi). \tag{3.48}$$

Now, observe that $h(\mathrm{x}|\phi)$ cannot be maximized in practice because it relies on some censored data. The idea of the EM algorithm is to maximize instead the average value of $\ln h(\mathrm{x}|\phi)$ knowing the observations z, *i.e.* $\mathrm{E}_{\mathrm{X}}[\ln h(\mathrm{X}|\phi)|\mathrm{z},\phi]$. It can be shown that the value ϕ maximizing this expectation is exactly the same as the one maximizing $g(\mathrm{z}|\phi)$ (this is a classical result of the EM algorithm). The steps of the algorithm can finally be written

1. At iteration $p+1$ compute $Q(\phi|\phi^{(p)}) := \mathrm{E}_{\mathrm{X}}[\ln h(\mathrm{X}|\phi)|\mathrm{z},\phi^{(p)}]$;
2. Find $\phi^{(p+1)} = \max_{\phi} Q(\phi|\phi^{(p)})$.

3.3.2 Computation of $Q(\phi|\phi^{(p)})$

To compute $Q(\phi|\phi^{(p)})$ we separate the product (3.48) into two terms, one including the known samples (NS) and the other considering the saturated samples (S).

$$\begin{aligned} Q(\phi|\phi^{(p)}) &= \mathrm{E}_{\mathrm{X}}[\ln h(\mathrm{X}|\phi)|\mathrm{z},\phi^{(p)}] \\ &\overset{(a)}{=} \sum_{j\in NS} \ln p(\mathrm{z}_j|\phi_j) + \sum_{j\in S} \mathrm{E}_{\mathrm{Z}_j}\left[\ln p(\mathrm{Z}_j|\phi_j)|\mathrm{z},\phi^{(p)}\right], \end{aligned} \tag{3.49}$$

(3.50)

(a) Since z_j is known for $j \in NS$, the value $\ln p(\mathrm{z}_j|\phi_j)$ is no longer a random variable but a deterministic value.

For $j \in S$ we do not know Z_j exact value but we know its distribution

$$\mathrm{E}_{\mathrm{Z}_j}\left[\ln p(\mathrm{Z}_j|\phi_j)|\mathrm{z}, \phi^{(p)}\right] = \int_{z_{sat}}^{\infty} \ln p(\mathrm{z}_j|\phi_j) \left(\frac{p(\mathrm{z}_j|\phi_j^{(p)})}{P_j(\phi_j^{(p)})}\right) d\mathrm{z}_j \tag{3.51}$$

$$= -\frac{1}{2}\left[\ln(2\pi\sigma_j^2) + \frac{\mu_j^2}{\sigma_j^2} + \frac{1}{\sigma_j^2}\int_{z_{sat}}^{\infty} \mathrm{z}_j^2 \left(\frac{p(\mathrm{z}_j|\phi_j^{(p)})}{P_j(\phi_j^{(p)})}\right) d\mathrm{z}_j \ldots \right.$$

$$\left. -\frac{2\mu_j}{\sigma_j^2}\int_{z_{sat}}^{\infty} \mathrm{z}_j \left(\frac{p(\mathrm{z}_j|\phi_j^{(p)})}{P_j(\phi_j^{(p)})}\right) d\mathrm{z}_j\right]. \tag{3.52}$$

We now compute

$$\int_{z_{sat}}^{\infty} \mathrm{z}_j^2 \left(\frac{p(\mathrm{z}_j|\phi_j^{(p)})}{P_j(\phi_j^{(p)})}\right) d\mathrm{z}_j \overset{(a)}{=} \mathrm{E}[\mathrm{Z}_j^2] \tag{3.53}$$

$$= \sigma_j^{2(p)}[1-\delta(\alpha)] + [\mu_j^{(p)} + \sigma_j^{(p)}\lambda(\alpha)]^2, \tag{3.54}$$

$$\tag{3.55}$$

(a) Z_j following the truncated normal distribution $\mathcal{N}(\mu_j^{(p)}, \sigma_j^{2(p)})$ in the interval $[z_{sat}, \infty)$. Hence, with

$$\alpha = \frac{z_{sat} - \mu_j^{(p)}}{\sigma_j^{(p)}} \quad \lambda(\alpha) = \frac{\phi(\alpha)}{1-\Phi(\alpha)} \quad \delta(\alpha) = \lambda(\alpha)(\lambda(\alpha) - \alpha), \tag{3.56}$$

$\phi(\cdot)$ is the probability density function of the standard normal distribution and $\Phi(\cdot)$ is its cumulative distribution function .
We continue with

$$\int_{z_{sat}}^{\infty} \mathrm{z}_j \left(\frac{p(\mathrm{z}_j|\phi_j^{(p)})}{P_j(\phi_j^{(p)})}\right) d\mathrm{z}_j \overset{(b)}{=} \mathrm{E}[\mathrm{Z}_j] \tag{3.57}$$

$$= \mu_j^{(p)} + \sigma_j^{(p)}\lambda(\alpha), \tag{3.58}$$

(b) Z_j following the truncated normal distribution $\mathcal{N}(\mu_j^{(p)}, \sigma_j^{2(p)})$ in the interval $[z_{sat}, \infty)$. Hence,

$$\mathrm{E}_{\mathrm{Z}_j}\left[\ln p(\mathrm{Z}_j|\phi_j)|\mathrm{z}, \phi^{(p)}\right] = -\frac{1}{2}\left[\ln(2\pi\sigma_j^2) + \frac{1}{\sigma_j^2}\left\{\mu_j^2 - 2(\mu_j^{(p)} + \sigma_j^{(p)}\lambda(\alpha))\mu_j + \ldots \right.\right.$$

$$\left.\left. +\sigma_j^{2(p)}(1-\delta(\alpha)) + (\mu_j^{(p)} + \sigma_j^{(p)}\lambda(\alpha))^2\right\}\right]. \tag{3.59}$$

Thus, from (3.50) and (3.59)

$$Q(\phi|\phi^{(p)}) = \sum_{j\in NS} \ln p(\mathrm{z}_j|\phi_j) - \frac{1}{2}\sum_{j\in S}\left[\ln(2\pi\sigma_j^2) + \frac{1}{\sigma_j^2}\left\{\mu_j^2 - 2(\mu_j^{(p)} + \sigma_j^{(p)}\lambda(\alpha))\mu_j + \ldots\right.\right. \tag{3.60}$$

$$\left.\left. +\sigma_j^{2(p)}(1-\delta(\alpha)) + (\mu_j^{(p)} + \sigma_j^{(p)}\lambda(\alpha))^2\right\}\right]. \tag{3.61}$$

3.3.3 The modified log-likelihood

The logarithm of the likelihood function $g(\mathrm{z}|\phi)$ is given by the sum of two terms,

$$\ln g(\mathrm{z}|\phi) = \underbrace{\sum_{j:k_j=1} \ln p_j(\mathrm{z}_j;\phi)}_{\text{term 1}} + \underbrace{\sum_{j:k_j=0} \ln P_j(\phi)}_{\text{term 2}} . \tag{3.62}$$

The first term is the sum of the log-likelihood function evaluated at the non-saturated samples, i.e. the log-likelihood function of the MLE approach discarding the saturated samples (classical MLE approach). The second term is the sum of the logarithm of the probabilities of the saturated samples to be saturated. It can be written as

$$\sum_{j:k_j=0} \ln P_j(\phi) = \sum_{j:k_j=0} \ln \left[\int_{z_{sat}}^{\infty} \frac{1}{\sqrt{2\pi\sigma_j^2(\phi)}} \exp\left\{ -\frac{1}{2\sigma_j^2(\phi)} (\mathrm{z} - \mu_j(\phi))^2 \right\} d\mathrm{z} \right] , \tag{3.63}$$

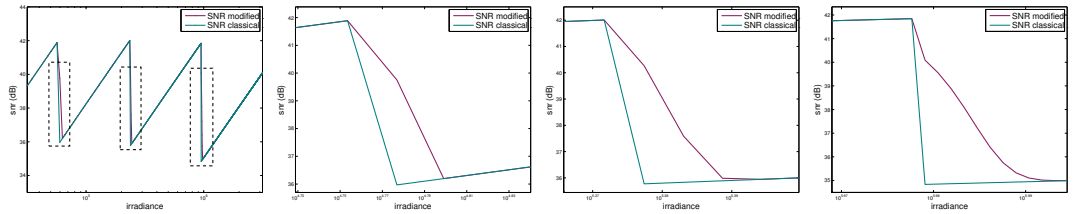
where $\mu_j(\phi) = ga\tau_j\phi$ and $\sigma_j^2(\phi) = g^2 a\tau_j\phi + \sigma_R^2$. Equation 3.63 is an increasing function that converges to zero for large ϕ, since large irradiance values have high probability to saturate. When we add the term (3.63) to the classical log-likelihood, its maximum will either: a) be almost unchanged, if it was reached on a value where (3.63) is almost zero; b) otherwise, be reached at a larger value.

Figure 3.9 shows an example of this modified log-likelihood $\ln g(\mathrm{z}|\phi)$ (red), the classical log-likelihood (blue), and the addition function (3.63) (green). In this example, the classical MLE is reached at a value ϕ where (3.63) is clearly negative. Therefore, when adding (3.63) and the classical log-likelihood, the maximum of the sum is reached for a larger ϕ. This example can be interpreted in the following way: if we use only the non-saturated samples we obtain the classical estimator $\hat{C}_{\text{MLE}}$. However, knowing that we have saturated samples we increase $\hat{C}_{\text{MLE}}$, since according to (3.63), the value $\hat{C}_{\text{MLE}}$ is not sufficiently likely to saturate for the saturated exposure times. A higher value is representative of both, the non-saturated samples through $\hat{C}_{\text{MLE}}$ and the saturated samples through the bias introduced by (3.63). Hence, the term (3.63) is adding a positive bias to $\hat{C}_{\text{MLE}}$ when needed.

Irradiance values close to the minimum value that saturates for a given exposure may saturate due to noise. In those cases, incrementing the estimation with the saturation bias may degrade the result if the classical MLE was already accurate. Nevertheless, it is seen that these cases rarely occur in practice and in average adding the bias factor always improves the results.

3.3.4 Modified Cramér-Rao Lower Bound

As performed in Section 3.2.2, the Cramér-Rao lower bound can be computed assuming the samples follow the modified log-likelihood function (3.62).



(a) All irradiance range. (b) Zoom 1st sat. level. (c) Zoom 2nd sat. level. (d) Zoom 3rd sat. level.

Figure 3.10: Comparison of the SNR obtained with the classical and the modified log-likelihoods. Both curves match except for the irradiance values in the transition between two saturation levels.

Proposition 5. *Given* $\mathrm{Z}_1, \ldots, \mathrm{Z}_T$*, independent Gaussian distributed random variables with mean and variance given by* $\mu_i = ga\tau_i C$ *and* $\sigma_i^2 = g^2 a\tau_i C + \sigma_R^2$ *(Model* (2.2)*), the Cramér-Rao lower bound for the estimation of the irradiance parameter* C *from* $\mathrm{z}_1, \ldots, \mathrm{z}_T$ *observations of these random variables assuming the modified probability density function* (3.62) *is given by*

$$\textsc{crlb}_{\textsc{mod}} = \frac{-1}{\sum_{j:k_j=1} E\left[\frac{\partial^2 \ln p(\mathrm{Z}_j;C)}{\partial C^2}\right] + \sum_{j:k_j=0} E\left[\frac{\partial^2 \ln P_j(C)}{\partial C^2}\right]}. \tag{3.64}$$

The first term in the denominator is the Fisher information for the classical log-likelihood (3.6)*. The second term can be computed as*

$$\sum_{j:k_j=0} E\left[\frac{\partial^2 \ln P_j(C)}{\partial C^2}\right] = \sum_{j:k_j=0} \frac{\int_{v_{sat}}^{\infty} \frac{\partial^2 A_j(C)}{\partial C^2} d\mathrm{z}}{\int_{v_{sat}}^{\infty} A_j(C) d\mathrm{z}} - \left[\frac{\int_{v_{sat}}^{\infty} \frac{\partial A_j(C)}{\partial C} d\mathrm{z}}{\int_{v_{sat}}^{\infty} A_j(C) d\mathrm{z}}\right]^2, \tag{3.65}$$

with

$$A_j(C) = \frac{1}{\sqrt{2\pi\sigma_j^2(C)}} \exp\left\{-\frac{1}{2\sigma_j^2(C)}(\mathrm{z} - \mu_j(C))^2\right\}. \tag{3.66}$$

Proof. The modified probability density function can be expressed as

$$\ln g(\mathrm{Z}|\phi) = \sum_{j:k_j=1} \ln p_j(\mathrm{Z}_j;\phi) + \sum_{j:k_j=0} \ln P_j(\phi). \tag{3.67}$$

The CRLB for the estimation of the parameter ϕ is computed as

$$\textsc{crlb} = \frac{-1}{E\left[\frac{\partial^2 \ln g(\mathrm{Z},\phi)}{\partial \phi^2}\right]} \tag{3.68}$$

$$= \frac{-1}{E\left[\frac{\partial^2 (\sum_{j:k_j=1} \ln p_j(\mathrm{Z}_j;\phi) + \sum_{j:k_j=0} \ln P_j(\phi))}{\partial \phi^2}\right]} \tag{3.69}$$

$$= \frac{-1}{\sum_{j:k_j=1} E\left[\frac{\partial^2 \ln p_j(\mathrm{Z}_j;\phi)}{\partial \phi^2}\right] + \sum_{j:k_j=0} E\left[\frac{\partial^2 \ln P_j(\phi)}{\partial \phi^2}\right]}. \tag{3.70}$$

Let us define

$$A_j(\phi, \mathrm{z}) := \frac{1}{\sqrt{2\pi\sigma_j^2(\phi)}} \exp\left\{-\frac{1}{2\sigma_j^2(\phi)}(\mathrm{z} - \mu_j(\phi))^2\right\}, \tag{3.71}$$

then we have

$$P_j(\phi) = \int_{v_{sat}}^{\infty} A_j(\phi, \mathrm{z}) d\mathrm{z}. \tag{3.72}$$

The first derivative of $\ln P_j(\phi)$ is given by

$$\frac{\partial \ln P_j(\phi)}{\partial \phi} = \frac{1}{\int_{v_{sat}}^{\infty} A_j(\phi, \mathrm{z}) d\mathrm{z}} \frac{\partial(\int_{v_{sat}}^{\infty} A_j(\phi, \mathrm{z}) d\mathrm{z})}{\partial \phi} \tag{3.73}$$

$$= \frac{\int_{v_{sat}}^{\infty} \frac{\partial A_j}{\partial \phi} d\mathrm{z}}{\int_{v_{sat}}^{\infty} A_j(\phi, \mathrm{z}) d\mathrm{z}}, \tag{3.74}$$

since $A(\phi, \mathrm{z})$ meets the conditions of the Leibnitz rule. Thus the second derivative is given by

$$\frac{\partial^2 \ln P_j(\phi)}{\partial \phi^2} = \frac{\int_{v_{sat}}^{\infty} \frac{\partial^2 A_j(\phi, \mathrm{z})}{\partial \phi^2} d\mathrm{z}}{\int_{v_{sat}}^{\infty} A_j(\phi, \mathrm{z}) d\mathrm{z}} - \left[\frac{\int_{v_{sat}}^{\infty} \frac{\partial A_j(\phi, \mathrm{z})}{\partial \phi} d\mathrm{z}}{\int_{v_{sat}}^{\infty} A_j(\phi, \mathrm{z}) d\mathrm{z}}\right]^2. \tag{3.75}$$

Since $\frac{\partial^2 \ln P_j(\phi)}{\partial \phi^2}$ does not depend on z, we have $E\left[\frac{\partial^2 \ln P_j(\phi)}{\partial \phi^2}\right] = \frac{\partial^2 \ln P_j(\phi)}{\partial \phi^2}$, thus

$$\sum_{j:k_j=0} \mathrm{E}\left[\frac{\partial^2 \ln P_j(\phi)}{\partial \phi^2}\right] = \sum_{j:k_j=0} \frac{\int_{v_{sat}}^{\infty} \frac{\partial^2 A_j(\phi)}{\partial \phi^2} d\mathrm{z}}{\int_{v_{sat}}^{\infty} A_j(\phi) d\mathrm{z}} - \left[\frac{\int_{v_{sat}}^{\infty} \frac{\partial A_j(\phi)}{\partial \phi} d\mathrm{z}}{\int_{v_{sat}}^{\infty} A_j(\phi) d\mathrm{z}}\right]^2. \tag{3.76}$$

□

Both curves, $\mathrm{CRLB}_{\mathrm{SAT}}$ and the Cramér-Rao lower bound for the modified log-likelihood $\mathrm{CRLB}_{\mathrm{MOD}}$, match everywhere except in the transition zones between two saturation levels. We name saturation level n the irradiance range for which the longest n exposures produce saturated samples. This was expected since the contribution of the modified log-likelihood to the $\mathrm{CRLB}_{\mathrm{SAT}}$ can be thought of as a measure of the information carried by the saturated samples. The higher information is in the region of higher uncertainty on whether pixels saturate or not. If we take the middle of the irradiance range between transition zones n and $n+1$, the corresponding pixels are highly likely to saturate for the longest n exposures and very unlikely to saturate for exposure $n+1$. Thus the information provided by the fact that the pixel saturates for the longest n exposures is not significant. The same behavior is found for the SNR curves and is illustrated in Figure 3.10. The curves differ in the transition zones only. Yet we observe that the estimation performance can be considerably increased for the concerned irradiance ranges.

3.3.5 Experiments

Synthetic data generation Synthetic samples are generated from a HDR image taken as ground-truth assuming the Model (2.2) for pixel values. The simulated camera is Camera A (c.f. Section 3.2.4). The exposure times are $\tau = (1/4.2, 1/16.8, 1/67.2, 1/268.8)s$.

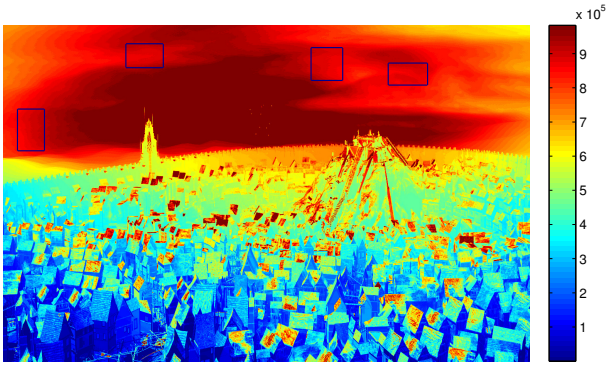


Figure 3.11: Ground-truth image. From [Media] under Creative Commons License. The results for the four marked sub-regions are shown in Figure 3.12.

Results Figure 3.12 shows the results for four sub-regions of the ground-truth image, indicated on Figure 3.11. The pixels in these sub-regions have 2 or 3 saturated samples. In all cases, the information provided by the saturated samples improves the result. The improvement can also be verified in Table 3.4 where the values of PSNR for each sub-region are presented.

3.4 Model parameters uncertainties and performance bounds

The results presented in Section 3.2.4 show that if the data follows Model (2.2), the MLE performs extremely well in estimating the irradiance. The estimation bias is negligible, and its variance gets very close to the CRLB_{SAT}. However, up to now we have assumed that the parameters that govern Model (2.2) are perfectly known. For practical purposes, this is of course not a realistic assumption: the gain factor, the readout noise mean and variance and the PRNU factors are unknown and have to be determined by means of a calibration procedure. Hence the model parameters are subject to uncertainties, whose impact in the irradiance estimation has to be quantified. The first part of this section is devoted to present an experimental study to assess how sensitive the MLE estimation is to variations in the model parameters.

In the second part of this section we concentrate on the consequences of ignoring the model parameter uncertainties when evaluating the performance of HDR generation techniques. Surprisingly, up to our knowledge, no previous work on HDR generation takes these uncertainties into account. Uncertainties in model parameters play a fundamental role in the creation of a ground-truth image from real, regular digital camera images. Hence, depending on how the ground-truth is generated, the impact of not considering model parameters uncertainties on the reported performance may vary. In any case, as we will see, ignoring these uncertainties may generally lead to overrated performances.

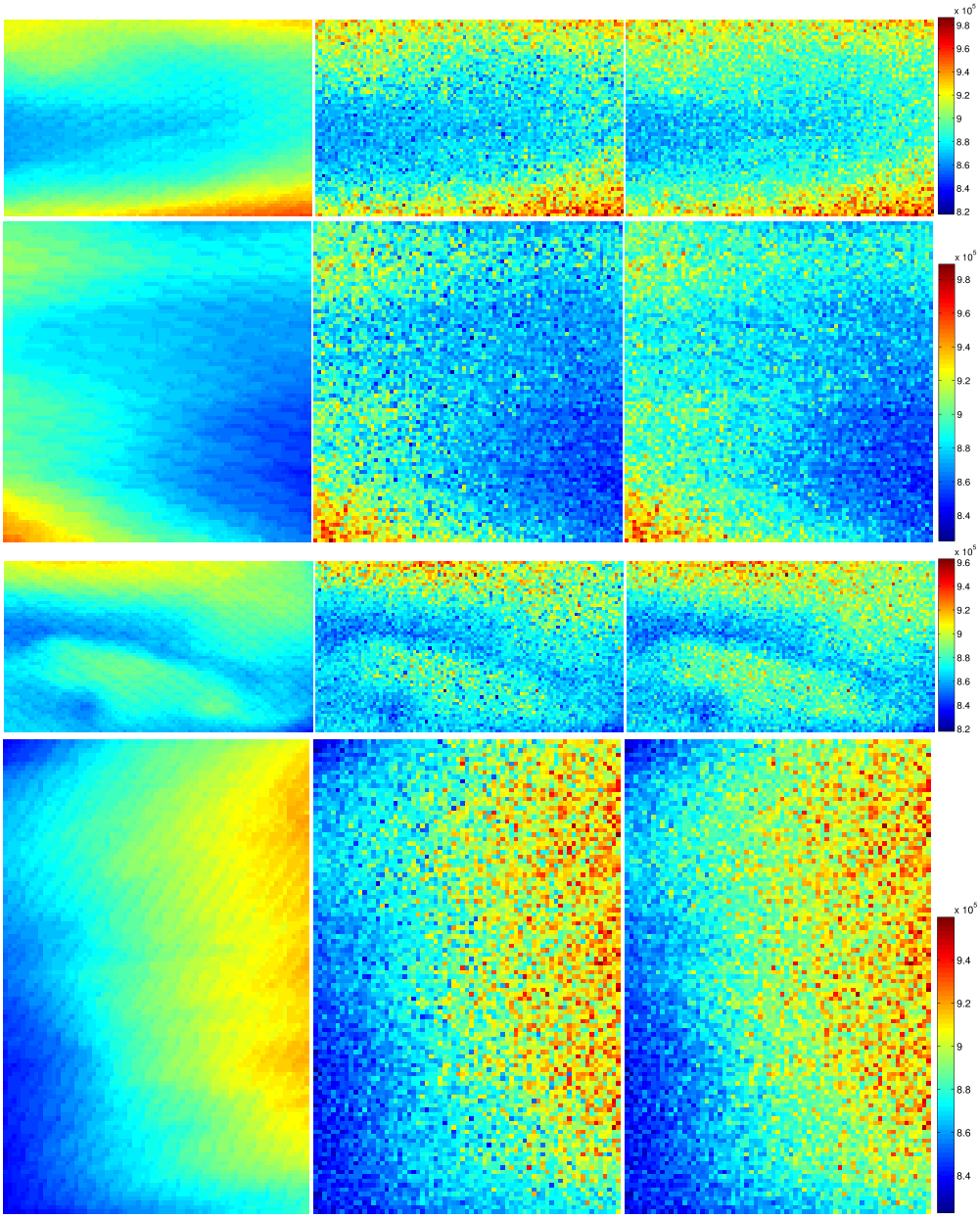


Figure 3.12: Results for four sub-regions of the ground-truth image in Figure 3.12. The pixels in these sub-regions have 2 or 3 saturated samples. Patch sizes from top to bottom 61×96, 84×81, 56×101, 107×69. **First column:** ground-truth. **Second column:** result obtained with the classical log-likelihood. **Third column:** result obtained with the modified log-likelihood. In all cases, the results are improved by the information provided by the saturated samples.

	PSNR (dB)			
	REGION 1	REGION 2	REGION 3	REGION 4
MODIFIED LOG-LIKELIHOOD	**37.7**	**38.7**	**39.3**	**38.1**
CLASSICAL LOG-LIKELIHOOD	36.9	37.6	38.3	37.0

Table 3.4: PSNR for the classical and modified likelihood for the four sub-regions marked in Figure 3.11.

Image size	g	σ_R^2	μ_R	a (n = 1)	a (n = 10)	a (n = 100)
100×100	0.013	0.447	0.056	0.013	0.004	0.001
200×200	0.006	0.224	0.028	0.013	0.004	0.001
500×500	0.003	0.089	0.011	0.013	0.004	0.001
1000×1000	0.001	0.045	0.006	0.013	0.004	0.001

Table 3.5: Standard deviation of the camera parameters estimated according to the procedure presented by Granados et al. [2010]. Real values are those of Camera A ($g = 0.87, \sigma_R^2 = 31.6, \mu_R = 2046$).

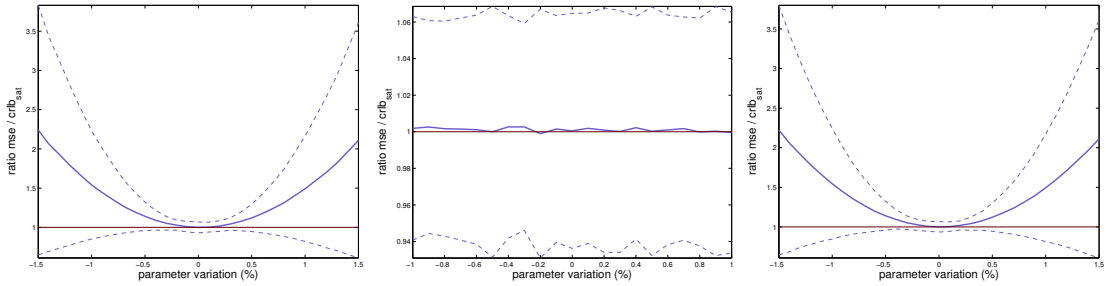
3.4.1 Sensitivity of the MLE to variations in the model parameters

The sensitivity analysis can only be performed with simulated images, since knowledge of the real parameter values is required. Synthetic data was generated according to Model (2.2), taking as ground-truth the HDR image shown in Figure 3.1 and the set of exposure times τ_{M6}. The two sets of camera parameters presented in Section 3.2.4 (Cameras A and B) were tested. These parameters are considered the real model parameters and used to simulate the data. Then they are varied and the maximum likelihood estimation is performed using the *wrong* parameters. The ratio between the MSE and the CRLB_{SAT} is computed in order to compare the results with those obtained using the exact model parameters.

The first step in this study is to establish realistic ranges for the uncertainties in the model parameters, which is given by the variance of the estimators used to find each parameter. The model parameters are estimated at the camera calibration stage and the variance of the estimators clearly depends on the calibration procedure. To do so, we consider the calibration procedure proposed by Granados et al. [2010]. They propose to compute the spatial mean and variance of a bias frame (a frame acquired with the cap on and with the shortest exposure) to find the mean and variance of the readout noise respectively. The gain and PRNU factors are computed using flat frames (frames acquired with uniform illumination). Using this calibration method we compute the variance of each model parameter.

Table 3.5 shows the standard deviation for the gain, the readout noise variance, the offset and the PRNU factors. The leftmost column represents the image size used to estimate the parameters. Several images are needed to estimate the PRNU factors, variable n represents the number images considered in each case. The real values of the parameters are those of camera A ($g = 0.87, \sigma_R^2 = 31.6, \mu_R = 2046$). The PRNU factors take Gaussian distributed values of mean 1 and variance 0.01. Similar results were obtained for Camera B.

The worst case standard deviations in Table 3.5 are used to define the variation range of the model parameters, namely: for g 1.5% of its value; 1.4% for σ_R^2; 2.7e-3% for μ_R and 1.3% for the PRNU factors. Because the variance of the estimator of the μ_R parameter is so small, the influence of the uncertainty on this parameter is not analyzed. Thus, the following variation ranges are considered for g, σ_R^2 and PRNU factors, respectively: $[0.985g, 1.015g]$ in steps of $0.001g$; $[0.99\sigma_R^2, 1.01\sigma_R^2]$, in steps of $0.001\sigma_R^2$; $[0.985a, 1.015a]$ in steps of $0.001a$ The influence of parameters uncertainties is analyzed individually: one of them is varied in its corresponding range, while the rest are kept fixed at their real values. For each tested (*wrong*) value, the estimation is repeated 1000 times for each irradiance



(a) g variation (b) σ_R^2 variation (c) PRNU variation

Figure 3.13: Dependence of the performance of the MLE estimator on the uncertainty in model parameters for Camera A (the results for Camera B are equivalent). A great dependence is found on both g and a. On the contrary, the variation for σ_R^2 is negligible.

level in order to compute the MSE.

Figure 3.13 shows the results for Camera A (the results for Camera B are equivalent). For each model parameter, the mean ratio of the MSE and the CRLB_{SAT} (blue) and the band of $\pm$ one standard deviation (dotted line) are shown as functions of the parameter deviation. The unity level is shown for reference.

As expected the curves of mean ratio reach 1 when using the real parameter value (zero variation). A strong impact on the MLE is observed for both g and PRNU factor (this similar behavior is reasonable since they play a very similar role in the model). In both cases, for small deviations (1.5% of the real value) the MSE of the MLE almost doubles the CRLB_{SAT}. Regarding the readout noise variance σ_R^2, the result of the MLE is not affected by variations in the considered range.

3.4.2 Model parameters uncertainties and performance evaluation

In order to evaluate the performance of a HDR image generation technique with real images, it is necessary to define a procedure to generate the ground-truth image. If the ground-truth is to be built from images taken with a regular digital camera (not with cameras able to capture directly HDR images), a HDR image generation technique has to be chosen to generate the ground-truth image. It turns out that the precision of the model parameters estimates plays a key role in this ground-truth image generation process.

A simple way to generate the ground-truth is to define a set of exposure times, take several images for each exposure and take for each pixel the average value corresponding to the best non saturated exposure. This ground-truth is clearly unbiased and for a large enough number of images per exposure time, the noise should be considerably reduced. Instead of taking the best exposure only, Granados et al. [2010] propose to combine the averaged images using the MLE with the weights computed from the variance of the images. If the model parameters are correctly estimated, both ground-truths are quite accurate. Nonetheless, if the parameters are not accurate these ground-truths will be strongly biased. Moreover, it is not difficult to show that if the irradiance estimation is carried out using the same parameters as the ground-truth generation, the bias present in both computations will partially compensate and the result will seem better than what it really is.

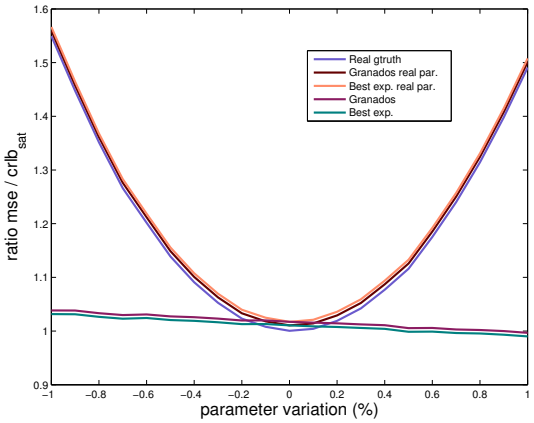


Figure 3.14: Dependence of the ground-truth computation with the uncertainty in model parameters. For small variations of the parameter the results of all computed ground-truths are almost equivalent. Yet for variations above 0.2% the performance announced by both ground-truths computed with the *wrong* parameters is highly superior than the real performance.

A synthetic experiment is performed to show this effect. Synthetic samples are generated as in the previous experiment, using Camera A configuration as real parameters and τ_{6L}. The gain spans the range $[0.985g, 1.015g]$ and the MLE is computed with the *wrong* parameters. Five different ground-truths are considered to compute the MSE: the real ground-truth (the image used to synthesize the samples), Granados's ground-truth obtained with the real parameters and with wrong parameters, the best exposure only ground-truth obtained with the real parameters and with wrong parameters. Figure 3.14 shows the ratio between the MSE of the MLE and the CRLB_{SAT} as a function of the variation in the gain. Each curve shows the results for a different ground-truth. For small variations of g the result given for the computed ground-truths is quite accurate. The ratio for the real ground-truth is below the other ratios, meaning that the result is actually better than claimed by the computed ground-truths. On the contrary, for deviations of g above 0.2%, the performance announced by both ground-truths computed with the *wrong* parameters is highly superior than the real performance.

3.5 Conclusions

In this chapter we have presented a study of the performance bounds of the HDR estimation problem, and we have analyzed the performance of current state-of-the-art estimation methods. This study shows that, to a first order approximation, the MLE is efficient even when using a very reduced number of samples. While its value can be computed numerically by iterative procedures, an approximation of the MLE cannot be directly derived since no closed-form exists. However, we observe that replacing the variance of each sample by its empirical value, yields to a closed-form which is extremely close to the MLE. Then we show that the first order Taylor expansion of the variance of this closed-form estimator, exhibits a negligible difference with the Cramér-Rao bound, for all irradiance values and for any number of samples. This result explains why, as previously claimed

based on experimental evidence [Granados et al. 2010], the MLE outperforms other estimation methods. In Chapter 5, the near optimality of the MLE is used to choose the irradiance estimator for a patch-based HDR imaging approach proposed for the case of dynamic scenes and hand-held camera.

Moreover, we have proposed a method that differs from all methods in the literature, in the sense that it integrates the information provided by saturated samples in the estimation process. The proposed approach follows closely the EM algorithm, in the version that considers censored data. Results confirm that saturated samples carry useful information for irradiance estimation, that allows to improve the irradiance estimation near the saturation values, without degrading the estimation in other irradiance ranges. Improvements in the order of 1 dB are found in experimental examples.

Finally, we have raised a delicate point that had not been addressed in previous studies on HDR estimation. We have shown that small errors in the calibration of camera parameters may severely degrade the estimation. In particular, when working with real data, a very accurate camera calibration is needed in order to obtain a reliable ground-truth. Otherwise, results may appear much better than what they are in reality. A natural continuation of this work is to derive irradiance estimators that are more robust to uncertainties on the camera parameters.

4 Simultaneous HDR and super-resolution

Not only the acquisition of the dynamic range of the scenes is limited when using conventional cameras, but also their frequency content. Even if the resolution of camera sensors have had an enormous increase in recent years, it remains limited by several physical factors. The need to go beyond these limits finds an answer in super-resolution (SR) methods, which aim at increasing the acquired frequency content by combining multiple images. The difference between sampling grids caused by the motion of the camera is used by super-resolution methods to generate a high resolution image. Considered as multi-imaging approaches, HDR imaging and super-resolution, can be combined to generate a super-resolved image with a full dynamic range.

In this chapter, we study the reconstruction error of high dynamic range super-resolution imaging without regularization. From this study, we deduce a strategy for the choice of the number of images and the exposure times which makes the unregularized problem well conditioned. With these acquisition parameters, if an affine motion hypothesis holds and sufficiently long exposure time is available, we show that the recovery of all the amplitude and frequency content of the scene irradiance is guaranteed. This work presented in this chapter is the result of a collaboration with Yann Traonmilin and has been accepted for publication at the SIAM Journal on Imaging Sciences. A pre-print of the article is available [Traonmilin and Aguerrebere 2014].

Super-resolution algorithms have been reviewed in several works [Farsiu et al. 2004; Milanfar 2010; Tian and Ma 2011]. Most of them can be summarized with a variational approach. The high resolution (HR) image is recovered by minimizing a regularized data-fit functional. This data-fit is usually an L^p norm fit. From Tychonov to total variation, many regularization functionals are available. Such regularization is necessary when little information is available (i.e. a small number of images). The resulting interpolation must be considered as an inpainting result using a regularity model and not as the recovery of missing high frequency information as the scene might not verify the regularity hypothesis. It has been shown [Champagnat et al. 2009; Traonmilin et al. 2012a] that regularization becomes less useful when a large number of images are available. It is then possible to reconstruct the real high frequency content of the acquired scene by minimizing only the L^2-norm data-fit.

As presented in Chapter 3, several HDR image generation algorithms have been proposed since the seminal work by Mann and Picard [1995]. For static scenes, when the input images can be perfectly registered, the irradiance at each pixel is computed as a weighted

average of the corresponding samples acquired with different exposure times or gains. Based on this idea, several methods have been proposed following different weighting schemes [Debevec and Malik 1997; Mitsunaga and Nayar 1999; Reinhard et al. 2005; Kirk and Andersen 2006; Granados et al. 2010; Hasinoff et al. 2010; Aguerrebere et al. 2014b]. These schemes have been studied in detail in Chapter 3. However, accurate registration of aliased images is sometimes impossible and a joint HDR-SR method is thus needed.

In recent years, some algorithms have been proposed to perform SR and HDR simultaneously. The most conventional approach is to use a weighted least squares scheme, as normally done for SR, but on the irradiance [Gunturk and Gevrekci 2006; Choi et al. 2009] or transformed irradiance domain [Bengtsson et al. 2012] instead of working directly with gray levels. A different approach [Zimmer et al. 2011] is to align low resolution (LR) images using optical flow and minimize a regularized energy to achieve the combined HDR-SR reconstruction. Ajdari Rad et al. [2007] propose to first align the input images and then use a Delaunay triangulation and bicubic interpolation to perform the HDR-SR estimation on the irradiance domain. The combined problem has also been studied from a sensor design perspective, with proposed solutions based on specifically adapted sensors [Narasimhan and Nayar 2005; Nakai et al. 2008].

An aspect common to all these methods is the use of a regularization term, which threatens multi-image super-resolution in terms of high frequency recovery capability. The recovery of the spectrum of the real HR scene cannot be guaranteed. Moreover, the latest results for HDR image generation, such as the near optimality of the weighting scheme proposed by Granados et al., are not exploited [Aguerrebere et al. 2014b]. To the best of our knowledge, the theoretical background for this problem, necessary to ensure the simultaneous recovery of HDR and HR content, has not been studied in depth.

In this work, we study the HDR-SR imaging problem from the following perspective: which image acquisition configuration, i.e. number of acquired images and exposure times, guarantees that the high frequencies and the full dynamic range of a scene will be recovered? We use a realistic camera acquisition model with affine motion hypothesis and propose an image reconstruction method that includes the weighting scheme shown to be nearly optimal for HDR information recovery [Aguerrebere et al. 2014b]. We study theoretical bounds for the joint HDR-SR reconstruction error and find that a trade off must be made between the total exposure time and the number of images. This result is illustrated with synthetic and real data experiments. Moreover, we show how exposure times can be selected and also that this selection can be decoupled from the SR problem to fulfill our objective. This analysis leads to the proposal of an acquisition strategy which, if the affine motion hypothesis holds and sufficiently long exposure time is available, guarantees the recovery of the full dynamic range and high frequency content of the scene irradiance. Finally, experiments with real data show that our HDR-SR strategy manages to recover this information.

The chapter is organized as follows. Section 4.1 introduces the image acquisition model and the reconstruction method. In section 4.2, the strategy for the choice of exposure times and number of images is proposed. Section 4.3 presents the study of the reconstruction error bounds. In Section 4.4, we show how exposure times can be selected. An optimal acquisition strategy and its corresponding experimental validation are presented

in Section 4.5. Finally, conclusions are stated in Section 4.6.

4.1 HDR-SR acquisition and reconstruction

4.1.1 Acquisition model

We consider a monochromatic irradiance acquisition model with different exposure times $(\tau_i)_{i=1,\dots,N}$ for each raw low resolution image $\mathrm{Z}_i \in \mathbb{R}^{l\times l}$. The acquisition operator A is defined as

$$\begin{aligned} \mathrm{A}: \quad & \mathbb{R}^{Ml\times Ml} \to \mathbb{R}^{N\times(l\times l)} \\ & \mathrm{C} \to (\mathrm{Z}_i)_{i=1,\dots,N} = (\Omega_i \mathrm{G}_i \mathrm{S} \mathrm{Q}_i \mathrm{C})_{i=1,\dots,N}, \end{aligned} \tag{4.1}$$

where N is the number of LR images, $\mathrm{C} \in \mathbb{R}^{Ml\times Ml}$ is the HR irradiance of size $Ml \times Ml$, $\mathrm{Q}_i \in \mathbb{R}^{(Ml\times Ml)\times(Ml\times Ml)}$ are the affine motions associated with each LR image ($\mathrm{Q}_i\mathrm{C}$ is the irradiance reaching the sensor), $\mathrm{S} \in \mathbb{R}^{(l\times l)\times(Ml\times Ml)}$ is the sub-sampling by a factor M (M can be called the super-resolution zoom or factor) and $\mathrm{G}_i = g\tau_i \in \mathbb{R}^{(l\times l)\times(l\times l)}$, is the overall acquisition gain, including the camera gain g (without loss of generality, we take $g = 1$ to simplify expressions) and the exposure time τ_i for the i-th LR image. Several images may be acquired with the same exposure time. $\Omega_i \in \mathbb{R}^{(l\times l)\times(l\times l)}$ is a diagonal matrix taking value 1 if pixel j does not become saturated or under-exposed for the exposure time τ_i and zero otherwise. We assume $\Omega = (\Omega_i)_{i=1,\dots,N}$ to be full rank (every pixel is at least illuminated once).

This model is close to the one introduced in [Gunturk and Gevrekci 2006]. We use a simplified version of it in order to facilitate the study. Unlike [Gunturk and Gevrekci 2006], the camera response function is here considered to be linear, which is a realistic model for raw images. Our model does not take into account the point spread function of the camera, which can be deconvolved afterward if motion is small [Traonmilin et al. 2012b]. Since we consider a monochromatic acquisition model, the Bayer pattern is ignored. With our model, the production of a demosaicked color HR image without regularization, would take the form of a separate inversion of the 3 color channels. Only the motion due to the shift in the grid would have to be included in Q_i. Note that the green channel would have 2 times more LR images than red and blue channels.

The digital LR images $\mathrm{Z} = (\mathrm{Z}_i)_{i=1,\dots,N}$ are contaminated by additive noise N:

$$\mathrm{Z} = \mathrm{AC} + \mathrm{N}. \tag{4.2}$$

Considering the acquisition noise model introduced in Chapter 2, we take N to be a spatially varying Gaussian noise

$$\mathrm{N} = \mathrm{N}_\mathrm{C} + \mathrm{N}_R, \tag{4.3}$$

where N_C is a Gaussian noise with covariance matrix $\mathbf{\Sigma}_\mathrm{C}$ proportional to the irradiance ($\mathbf{\Sigma}_\mathrm{C} = diag(\mathrm{AC})$) and N_R is a Gaussian readout noise with constant variance σ_R^2. Thus the covariance matrix of N is $\mathbf{\Sigma} = diag(\mathrm{AC}) + \sigma_R^2\mathbf{I}$ ($\mathbf{I}$ = identity matrix). We do not take into account the spatially variable gain caused by the photo-response non uniformity of the sensor (see Chapter 2), which should be included for a more accurate modeling. The objective is to recover C from Z.

4.1.2 Image reconstruction

It has been shown in [Traonmilin et al. 2012a] that M^2 images are necessary to perfectly recover the HR image in the case of constant exposure times. Because we assume that the HDR acquisition model is a full rank linear map, the composition of the two processes is invertible with M^2 LR images. We minimize the L^2 data-fit to recover an estimate of C

$$\tilde{\mathrm{C}} = \mathrm{argmin}_{\mathrm{C}} \|\mathrm{W}^{1/2}(\mathrm{AC} - \mathrm{Z})\|_2^2. \tag{4.4}$$

Multiplying by $\mathrm{W}^{1/2} = \mathbf{\Sigma}^{-1/2}$ normalizes the noise to have constant variance, thus the solution to (4.4) gives the minimum variance linear unbiased estimator of the irradiance. Practically, this problem is solved with a linear conjugate gradient calculation of $(\mathrm{W}^{1/2}\mathrm{A})^{\dagger}\mathrm{W}^{1/2}\mathrm{Z} = (\mathrm{A}^H\mathrm{WA})^{-1}\mathrm{A}^H\mathrm{WZ}$.

Since the weights W depend on the irradiance C, an iterative procedure is needed to complete the estimation. The weights can be initialized with a smoothed version of the LR images. In practice, it is found that one iteration yields good results, and the irradiance remains almost unchanged after the first iteration. The iterative computation of this estimator for the HDR image generation in the case of perfectly registered images, is introduced by Granados et al. [2010] and shown to be nearly optimal in [Aguerrebere et al. 2014b] (Chapter 3).

An experiment is conducted to evaluate the benefit of using the weights W. For this purpose, synthetic samples are generated from a ground-truth HDR image (from [Hasinoff et al. 2010], see Figure 4.1) according to Model (4.2), and the reconstruction is performed solving (4.4), with $\mathrm{W}^{1/2} = \mathbf{\Sigma}^{-1/2}$ and $\mathrm{W} = \mathbf{I}$. The experiment is repeated 100 times, with randomly chosen affine transformations and noise realizations. The signal-to-noise ration (SNR) for each reconstructed pixel is computed from the ground-truth value and the mean square error (MSE) at that pixel (MSE computed from the 100 experiments). Figure 4.1 shows the SNR images in decibels for each weighting type. The results greatly improve when the weights $\mathrm{W}^{1/2} = \mathbf{\Sigma}^{-1/2}$ are included. The improvement (gain in SNR) is more remarkable in the darker regions of the scene where the input samples are noisier. An average gain of 3.9 dB (average SNR for all pixels) is found when comparing the cases with weights (32.5 dB) and without weights (28.6 dB).

For image registration we use the variable projection [Traonmilin et al. 2012b] method adapted to our joint HDR-SR reconstruction problem.

4.2 Strategy for the choice of acquisition parameters

In order to establish an HDR-SR reconstruction strategy, we must define the set of exposure times and the number of images (per exposure time) to acquire. Given a set of exposure times $\tau_1, \ldots, \tau_N$, the reconstruction error can be arbitrarily large because the condition number of the acquisition operator A cannot be bounded. For instance, it is infinite if the camera does not move ($\mathrm{Q}_i = \mathbf{I}$). Therefore, it is not possible to find the exposure times minimizing the reconstruction error for the HDR-SR combined problem without knowledge of the camera motion. In practice, the motion is seldom known. In particular, it is unknown for images acquired with a hand-held camera.

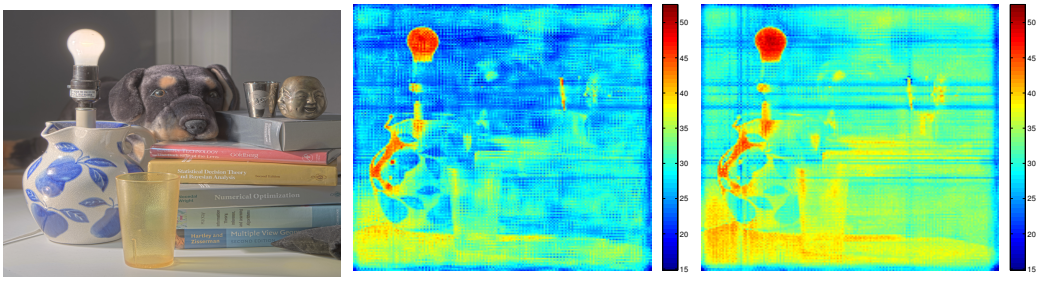


Figure 4.1: Result of optimal reconstruction for HDR-SR. SNR for each pixel obtained from 100 experiments with randomly chosen affine transformations and noise realizations. **Left:** Ground-truth image from [Hasinoff et al.]. **Center:** SNR obtained without weights ($\mathrm{W} = \mathbf{I}$). **Right:** SNR obtained using $\mathrm{W}^{1/2} = \mathbf{\Sigma}^{-1/2}$. The improvement is more remarkable in the darker regions of the scene where the input samples are noisier. An average gain of 3.9 dB (average SNR for all pixels) is found comparing the cases with weights (32.5 dB) and without weights (28.6 dB).

Performing a joint optimization of exposure times and number of images for HDR-SR would require a precise statistical model describing camera motion. Even with this information, calculating precise estimates of the reconstruction error would be difficult because we would have to estimate the interaction between the local conditioning of super-resolution [Traonmilin et al. 2012a] and the spatial variations of the SNR. Moreover, this joint problem is a highly dimensional non-convex problem, which would be hard to embed in a camera.

Hence, we propose to define a reconstruction strategy combining the optimality results for each separate problem, HDR and SR, so as to minimize the worst case reconstruction error of the combined problem. With HDR processing, we know that some parts of the scene are illuminated by few images. From SR theory, these parts must be covered by enough images to enable a reconstruction without regularization. Consequently, we proceed as follows.

Proposed strategy For the HDR reconstruction problem, an optimal HR HDR image is obtained if p HR images with exposures $\tau_1, \dots, \tau_p$ are combined, these exposures being chosen to minimize the HDR reconstruction error (as will be detailed in Section 4.4). Hence, we propose to consider the SR conditions which best estimate each of these p HR images. This will guarantee that the joint HDR-SR reconstruction performs well.

Suppose that we find a set of motions associated with each $\tau_1, \dots, \tau_p$ leading to a good SR reconstruction of the p HR images. Then a naive linear HDR-SR method would be to perform a classical HDR technique, with these p reconstructed HR images, thus limiting the final reconstruction noise. But we know that our joint method (minimization (4.4)) leads to the optimal L^2 reconstruction error, given this dataset. Consequently, by minimizing the reconstruction error of the separate HDR and SR problems, we guarantee a good reconstruction error for the joint HDR-SR problem.

Given a set of exposure times $\tau_1, \dots, \tau_p$ that minimize the HR HDR reconstruction error, two cases are to be considered. If the total exposure time and the total number of images are not limited, the best strategy for each exposure $(\tau_i)_{i=1,\dots,p}$ is to take as many LR images as possible with this exposure in order to improve the SR reconstruction. With

equal exposure times, the problem becomes a pure super-resolution problem where more images improve the quality of the reconstruction [Champagnat et al. 2009; Traonmilin et al. 2012a]. Longer exposures are not desirable since they increase the number of saturated pixels thus increasing the reconstruction error. A much more realistic case, is that of a limited total exposure time (e.g. because of scene motion). In order to ensure a correct HR reconstruction with a super-resolution factor (or zoom) M, the minimum total exposure time for frames acquired with time τ_i is $M^2\tau_i$. However, as it will be shown in the following section, there is an optimal number of LR images $N > M^2$ for a fixed total exposure time.

4.3 Reconstruction error bound for HDR-SR

In the following section we study the reconstruction error bounds for the HDR-SR estimation problem for a fixed total exposure time T and N images acquired with equal exposure T/N, i.e. the reconstruction of the HR irradiance frame.

4.3.1 Optimal number of images for a fixed total exposure time

In a HDR context, when neglecting motion blur, the longer the exposure time without saturation, the better. We show here that when we must perform SR at the same time (i.e. compute the pseudo-inverse of A), taking more images with a shorter exposure can be better. We study the case of HDR-SR for a given total exposure time T without saturation, *i.e.* $\Omega_i = \mathbf{I}$. N images are acquired with a total exposure time T, each with equal exposure time $\tau = T/N$. Because the overall acquisition gain is linear, the problem is equivalent to the acquisition of N LR irradiances Y with $\mathrm{A}' = \mathrm{G}^{-1}\mathrm{A}$, with a noise N' with covariance matrix:

$$\mathbf{\Sigma}' = diag(\mathrm{A}'\mathrm{C}/\tau + \sigma_R^2/\tau^2\mathbf{I}). \tag{4.5}$$

Then,

$$\mathrm{Y} = \mathrm{A}'\mathrm{C} + \mathrm{N}'. \tag{4.6}$$

A' is the conventional super-resolution modeling operator, *i.e.* a super-resolution operator where acquisition gain is not considered. Thanks to this normalization, we will be able to use results from the super-resolution literature directly.

The super-resolution of N images with exposure T/N can be done solving the problem

$$\tilde{\mathrm{C}} = \mathrm{argmin}_{\mathrm{Y}} \|\mathrm{W}'^{1/2}(\mathrm{A}'\mathrm{C} - \mathrm{Y})\|_2^2, \tag{4.7}$$

with $\mathrm{W}' = \mathbf{\Sigma}'^{-1}$.

Reconstruction error bound

Proposition 6. *The noise in the reconstructed image* N_{rec} *is thus bounded by*

$$\|\mathrm{N}_{rec}\|_2^2 \leq \frac{\kappa(N)^2}{\sigma_{max}^2(\mathrm{A}')} \frac{m^2 l^2}{\tilde{m}} (1 + rN) N^2, \tag{4.8}$$

where $\kappa(N)$ is the conditioning of $\mathrm{A'}^H\mathrm{A'}$ (ratio of extremal eigenvalues of $\mathrm{A'}^H\mathrm{A'}$), $\sigma_{max}(\mathrm{A'})$ is the maximum singular value of $\mathrm{A'}$, $m = sup(\mathrm{C})$, $\tilde{m} = inf(\mathrm{C})$, $r = \sigma_R^2/(Tm)$, $\tilde{r} = \sigma_R^2/(T\tilde{m})$ and l is the size of the input images. Hence, the optimal number of images for a fixed total time T is the one that minimizes $\|\mathrm{N}_{rec}\|^2$.

Proof. First, using operator norm inequalities:

$$\|\mathrm{N_{rec}}\|_2^2 = \|(\mathrm{A'}^H\mathrm{W'}\mathrm{A'})^{-1}\mathrm{A'}^H\mathrm{W'}\mathrm{N_{in}}\|_2^2 \tag{4.9}$$

$$\leq \|(\mathrm{A'}^H\mathrm{W'}\mathrm{A'})^{-1}\|_2^2\|\mathrm{A'}^H\mathrm{W'}^{1/2}\|_2^2\|\mathrm{W'}^{1/2}\mathrm{N_{in}}\|_2^2 \tag{4.10}$$

$$\leq \frac{\sigma_{max}^2(\mathrm{A'})\sigma_{max}^2(\mathrm{W'}^{1/2})}{\sigma_{min}^4(\mathrm{A'})\sigma_{min}^4(\mathrm{W'}^{1/2})}l^2N \tag{4.11}$$

where $\sigma_{min}(\mathrm{A'})$ and $\sigma_{max}(\mathrm{A'})$ are the minimum and maximum singular values of $\mathrm{A'}$ (respectively for $\mathrm{W'}$) and l is the size of input LR images. Let $\kappa(N)$ be the conditioning of $\mathrm{A'}^H\mathrm{A'}$, *i.e.* $\kappa(N) = \frac{\sigma_{max}^2(\mathrm{A'})}{\sigma_{min}^2(\mathrm{A'})}$. Using the fact that $\mathrm{W'}$ is diagonal and its definition from equation (4.5),

$$\|\mathrm{N_{rec}}\|_2^2 \leq \frac{\kappa(N)^2}{\sigma_{max}^2(\mathrm{A'})}\frac{(mN/T + \sigma_c^2N^2/T^2)^2}{\tilde{m}N/T + \sigma_c^2N^2/T^2}l^2N \tag{4.12}$$

$$\leq \frac{\kappa(N)^2}{\sigma_{max}^2(\mathrm{A'})}\frac{m^2l^2(1+rN)}{\tilde{m}(1+\tilde{r}N)}(1+rN)N^2 \tag{4.13}$$

where $m = sup(\mathrm{C}) \backsimeq sup(\mathrm{A'C})$, $\tilde{m} = inf(\mathrm{C}) \backsimeq inf(\mathrm{A'C})$, $r = \sigma_R^2/(Tm)$, $\tilde{r} = \sigma_R^2/(T\tilde{m})$. Finally, making use of the inequality $\frac{(1+rN)}{(1+\tilde{r}N)} \leq 1$ we have

$$\|\mathrm{N_{rec}}\|_2^2 \leq \frac{\kappa(N)^2}{\sigma_{max}^2(\mathrm{A'})}, \frac{m^2l^2}{\tilde{m}}(1+rN)N^2. \tag{4.14}$$

□

In [Traonmilin et al. 2012a], it was shown that $\kappa(N)$ can be bounded in probability by a decreasing function of the number of images N. Moreover, it can be shown [Traonmilin et al. 2013] that $N/M^2 \leq \|\mathrm{A'}\|_2^2 = \sigma_{max}^2(\mathrm{A'}) \leq N$. Hence, to optimize the reconstruction error bound, we minimize the function

$$f(N) = \kappa(N)^2(1+rN)N, \tag{4.15}$$

with respect to N. When N is close to the critical case M^2, $\kappa(N)$ can be large but decreases to 1 with N [Traonmilin et al. 2012a]. The conditioning of the SR operator has been also studied extensively in [Champagnat et al. 2009; Robinson and Milanfar 2006]. Consequently, f will have a minimum. Notice that the r factor is proportional to σ_R^2 and cannot be neglected in low light conditions. If we added saturation for a particular time T/N_0 with $N_0 \geq M^2$, the bound would still be valid for $N > N_0$. For $N \leq N_0$, matrix $\mathrm{A'}$ is not invertible, thus the error is not bounded.

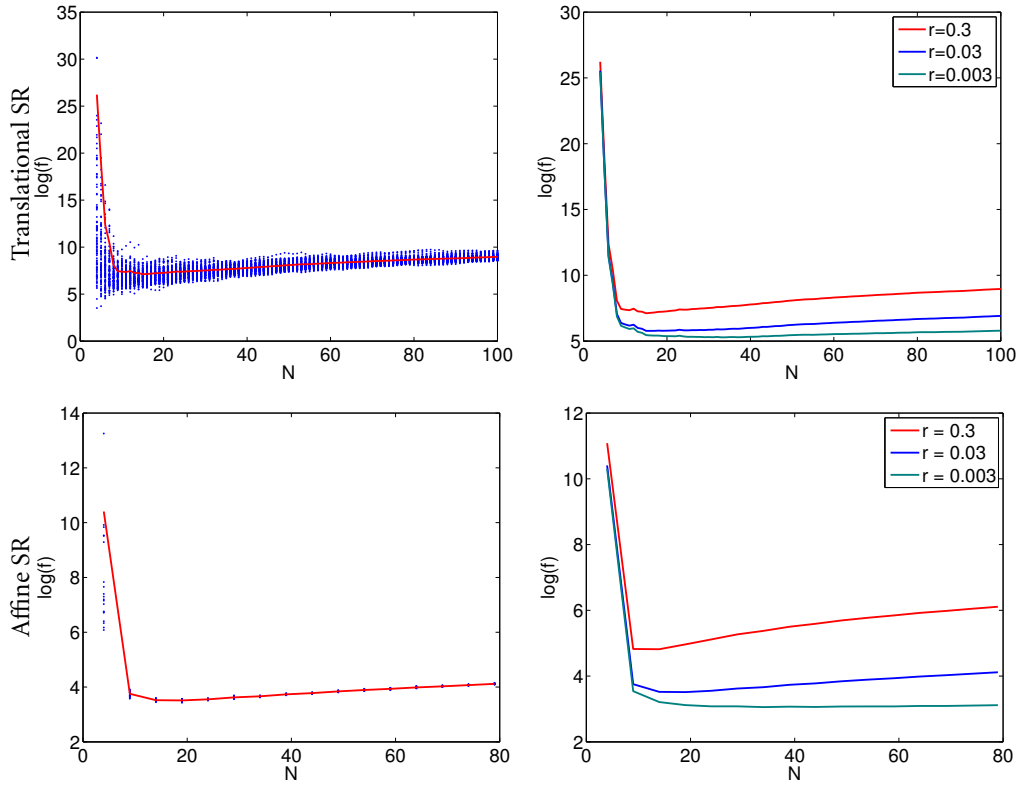
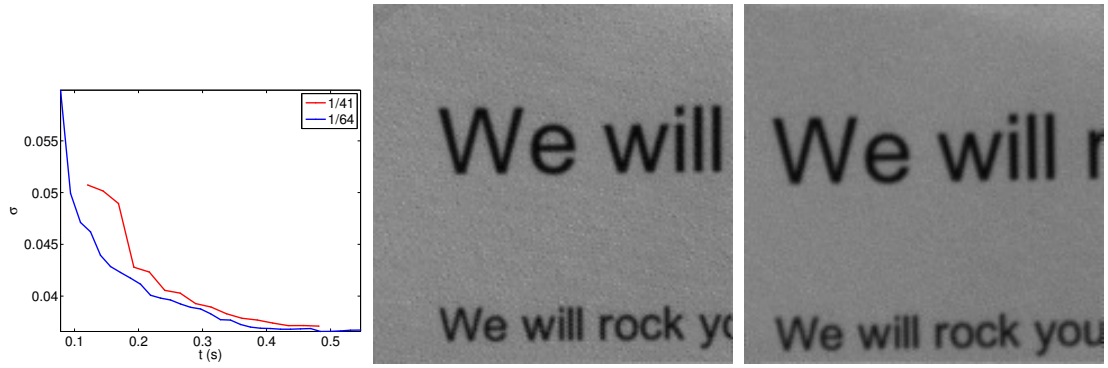


Figure 4.2: Plot of $\log(f(N))$ versus N ($M = 2$). **First row: (left)** Result for 100 experiments (blue) and average (red) of 2D translational SR for each N with $r = 0.03$. The minimum is reached at $N = 12$. **(right)** Plot of the average $\log(f)$ for different values of r for translational SR. Minimum at 12 (r=0.3), 13 (r=0.03) and 30 (r=0.003). **Second row: (left)** Affine super-resolution, 20 experiments for each N. The minimum is reached at $N = 19$, **(right)** Plot of average $\log(f(N))$ versus N for different values of r for affine SR. Minimum at 14 (r=0.3), 19 (r=0.03) and 34 (r=0.003).

Experimental study of $f(N)$

Synthetic data To illustrate the behavior of $f(N)$, we compute its curve for the case of 2D translational super-resolution. For each number of images N, we randomly simulate 100 translation parameters and compute the corresponding $f(N)$ by explicitly calculating $\kappa(N)$ (which can be done because the SR is then a 2D Vandermonde system in the frequency domain [Papoulis 1977; Ahuja and Bose 2006]). With the same procedure, we generate the $f(N)$ curve for the affine super-resolution case. Here $\kappa(N)$ is approximated by the ratio $\|\mathrm{N_{rec}}\|^2/\|\mathrm{N_{in}}\|^2$. Figure 4.2 shows the results obtained with $M = 2$ and $r = 0.03$ (chosen with realistic values: $\sigma_R^2 = 30$, $T = 1/10$, $m = 10^4$).

It can be verified that the minimum of f is not reached at the critical case ($N = M^2 = 4$), meaning that it is better to take more than M^2 images with shorter exposure times. This fact shows that a compromise must be made between the number of images needed to perform super-resolution (more than M^2) and the noise level on those images. Given the total exposure time T, a degradation of the performance can be observed for large N. This can be explained by the fact that each image has a shorter exposure time. For short enough exposures, the variance σ_R^2 of the readout noise becomes prominent. The



(a) Fixed total exposure time. (b) SR 7 images at 1/41s. (c) SR 10 images at 1/64s.

(d) Fixed number of images. (e) LR image at 1/41s. (f) LR image at 1/64s.

Figure 4.3: Reconstruction error with respect to total exposure time and number of images ($M = 2$). (a) Reconstruction error with respect to total time. (b) Reconstruction error with respect to number of images. (c) SR result with 7 images at $1/41$s. (d) One LR image with exposure $1/41$s. (e) SR result with 10 images at $1/64$s. (f) one LR image at with exposure $1/64$s.

amount of noise in LR images becomes too big to be compensated by the averaging effect of super-resolution, thus increasing the reconstruction error.

This can be verified with the results obtained for various r values in Figure 4.2. The r factor can be thought of as the inverse of the dynamic range of the acquired scene, since it is equal to the ratio between the constant noise variance σ_R^2 and the maximum acquired irradiance mT (up to the factor N). As the value of r decreases, the minimum of $f(N)$ is reached at a larger N. This is due to the fact that both, decreasing σ_R^2 or increasing T, improve image quality and thus allow the use of shorter exposure times. Moreover, notice that the error bound increases quite slowly after its minimum. Hence, whether we choose a number of images at the minimum or slightly above will have little impact on the results. Curves of similar shape are obtained for larger M values, with the minimum reached for a larger N. This is caused by the larger conditioning of the SR problem for larger M [Baker and Kanade 2002].

Real data In order to experimentally verify the previous results, we take pictures of a planar surface using a hand held camera (c.f Figure 4.3). The small hand motion gives the sampling diversity needed for HDR-SR and guarantee that motion blur is not too

large. The goal is to compare the HDR-SR reconstruction error ($M = 2$) obtained with a constant total exposure time T, but with two different exposure times, $\tau_1 = 1/41$ and $\tau_2 = 1/64$, *i.e.* with a different number of images, $N_1 = T/\tau_1$ and $N_2 = T/\tau_2$. In the same way, the HDR-SR results are compared for the two different exposure times when fixing the total number of images N, i.e. for a different total exposure time $N\tau_1$ and $N\tau_2$ in each case. The reconstruction error (estimated from the variance of gray parts) with respect to the total exposure time and the number of images, for the two different exposure times, is shown in figures 4.3 (a) and (d) respectively. In this example, for a fixed number of images, it is better to take pictures at $1/41$ seconds. However, for a fixed total exposure time, it is better to combine more images with the shorter exposure $1/64$ seconds. The same behavior can be observed in the extracts of the reconstructed images shown in figures 4.3 (b) and (c). Hence, we verify that the longest exposure time is not necessarily optimal.

4.3.2 Discussion and practical consequences

A trade-off must be made when performing HDR and SR simultaneously: in an HDR setting, we look for the biggest exposure times which do not saturate (or saturate for the fewest images possible) while with SR, more images are generally better. This leads to the following remarks depending on the acquisition setup:

- **With a fixed total exposure time**, the number of images must be above the critical case for SR (M^2), thus with a shorter exposure. However, it must not exceed by much the critical case since for very short exposures the reconstruction noise increases with N.
- **With a fixed number of images**, the more images with the longest possible exposure time the best. However, the previous remark give us the knowledge that if this number of images is the optimal one from the previous section, then it is not possible to do better with the same total exposure time.
- **With any exposure time / number of images**, the best strategy is taking as much images as possible of the longest exposure time. Taking a lot of images with very short times is not a good strategy.

In the following section we analyze the problem of the exposure times selection focusing on the extension of the dynamic range of the image ($M = 1, \mathrm{Q}_i = \mathbf{I}, \mathrm{S} = \mathbf{I}$).

4.4 Exposure times selection for HDR reconstruction

The selection of the exposure times is a key aspect of the HDR imaging problem. Several approaches can be found in the literature that tackle this problem from different perspectives. Some of the proposed methods look for the optimal times set that ensures that a predefined dynamic range will be captured [Grossberg and Nayar 2003; Barakat et al. 2008]. Another group of methods focuses on the quality of the reconstructed HDR image and optimizes a risk function that depends on the mean squared error [Hirakawa and

Wolfe 2010] or the signal to noise ratio of the reconstruction [Granados et al. 2010; Hasinoff et al. 2010]. These methods need a prior estimation of the scene irradiance, which is assumed to be available for instance through the irradiance histogram.

In this chapter, we focus on the minimization of the HDR-SR reconstruction error, and we seek for the exposure times set that minimizes the HDR reconstruction error for a given number of exposures N'. The determination of the number of different exposure times N' is an important problem in HDR imaging, since it has a major impact on the quality of the results and on the practical constraints of the acquisition. Given that it is not the goal of this chapter to study this problem, we consider instead the value N' as a given parameter, which might eventually be set from the known irradiance histogram. An histogram of the scene irradiance is assumed known. In this section we concentrate on the extension of the dynamic range of the image ($M = 1, \mathrm{Q}_i = \mathbf{I}, \mathrm{S} = \mathbf{I}$) and propose an algorithm to find this exposure times set.

In this case, the solution to equation (4.4) gives the following estimator of the irradiance at position j:

$$\tilde{\mathrm{C}}_j = \frac{\sum_{i=1}^{N'} \frac{\tau_i \Omega_{ij} \mathrm{Z}_{ij}}{\mathrm{C}_j t_i + \sigma_R^2}}{\sum_{h=1}^{N'} \frac{t_h^2 \Omega_{hj}}{\mathrm{C}_j t_h + \sigma_R^2}}. \tag{4.16}$$

Thus the irradiance estimator variance in this case is given by

$$\|\mathrm{N}_{\mathrm{rec}}\|_2^2 = \sum_{j=1}^{l^2 M^2} \frac{1}{\sum_{h=1}^{N'} \frac{t_h^2 \Omega_{hj}}{\mathrm{C}_j t_h + \sigma_R^2}}. \tag{4.17}$$

Then, we consider the optimal set of times as the one that minimizes (4.17).

If we neglect the readout noise variance σ_R^2, the irradiance estimator variance expressed in (4.17) depends on the square root of the signal $\sqrt{\mathrm{C}_j}$. Thus, the minimization of (4.17) prioritizes noise reduction in bright regions. We opt to normalize the estimator variance by the signal as a way to counter this dependence and prioritize noise reduction in low light pixels, since noise is generally amplified in those pixels by post-processing tone mapping techniques used to display HDR images. Several tone mapping techniques perform a contrast reduction of the irradiance image that magnifies the noise visibility mainly in dark regions [Reinhard et al. 2010]. Nevertheless, this is clearly one possible option among multiple valid options. Other approaches are also valid (e.g. normalization by the squared signal C_j^2 so as to minimize the squared signal to noise ratio) and may be incorporated into the proposed HDR-SR acquisition strategy by modifying (4.17) accordingly. Therefore, we consider the normalized irradiance estimator variance

$$\|\mathrm{N}'_{\mathrm{rec}}\|_2^2 = \sum_{j=1}^{l^2 M^2} \frac{\frac{1}{\mathrm{C}_j}}{\sum_{h=1}^{N'} \frac{t_h^2 \Omega_{hj}}{\mathrm{C}_j t_h + \sigma_R^2}} = \sum_{j=1}^{l^2 M^2} \frac{1}{\sum_{h=1}^{N'} \frac{t_h^2 \Omega_{hj}}{t_h + \sigma_R^2 / \mathrm{C}_j}}, \tag{4.18}$$

and the set of exposure times

$$\tilde{t}_1, \ldots, \tilde{t}_{N'} = \arg \min_{t_1, \ldots, t_{N'}} \|\mathrm{N}'_{\mathrm{rec}}\|_2^2. \tag{4.19}$$

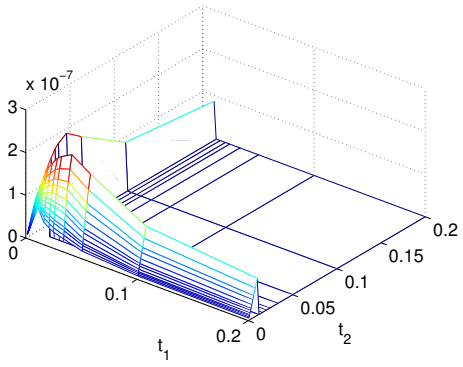


Figure 4.4: Inverse of the normalized reconstruction error (inverse of Equation (4.18)) for the first example scene of Section 4.5.

To minimize the non-convex function (4.18), we propose to use an exhaustive evaluation method. First, the shorter exposure time is determined by the fact that all irradiance values must be correctly acquired for at least one exposure time (matrix A is full rank). For the same reason, a lower bound is determined for the longer exposure. Then, given the number N' of exposure times to acquire (which might be automatically extracted from the scene histogram), we need to minimize function (4.18) with respect to the remaining $N' - 1$ times. We find this minimum by evaluating all combinations of available times. This methodology is feasible and fast in practice since N' is usually in the order of 2 to 4 and the number of available exposures in the camera is about 50 (e.g. 55 for the Canon 7D).

Figure 4.4 shows the inverse of function (4.18) for the example scene presented on the first experiment of Section 4.5 with $N' = 2$. The large flat region for the longer exposure times represents the large error caused by the saturation of most (or all) pixels in the image. A large error is also found for very short exposure times, since the readout noise becomes important for all irradiance values. The maximum of the function (minimum of (4.18)) is reached between these two extremes, as a compromise between saturation and noise level. This compromise is determined from the proportion of pixels belonging to each irradiance range, which is derived from the irradiance histogram of the scene.

4.5 Acquisition strategy

From the results presented in previous sections we derive the following strategy for the a regularized HDR-SR reconstruction that guarantees the recovery of the amplitude and frequency content of the scene irradiance. The exposure times are first selected following the procedure introduced in section 4.4. For this purpose, two images are taken to compute the histogram of the scene irradiance and find the optimal exposure times for the HDR reconstruction. One image must capture bright regions (no saturated pixels) and the other must capture dark regions (no under-exposed pixels). From the concatenated histograms of the two exposures, the high dynamic range histogram of the irradiance is built. With this histogram, the minimization from equation (4.19) is performed for the given number of exposure times.

Figure 4.5: Practical set-up of the first experiment.

The number of images to take for each exposure time is thus defined from the study presented in section 4.3. This number should be taken as the minimum of function f according to the second remark in section 4.3.2. The function f, thus the optimal N, depends on the scene irradiance. However, for $M = 2$, we find in practice that it is reached at about $N = 20$ for a large range of irradiances. It can be larger for smaller r values, but in those cases the function is quite flat and the result is not greatly affected by this choice. Because $N = 20$ is a feasible number of shots for burst acquisition with a hand-held camera, it can be taken as the reference number of images needed for each exposure time. Hence, 20 hand-held pictures are taken for each of the selected exposures and the joint HDR-SR is performed solving (4.4) with all images. As stated in Section 4.2, the joint HDR-SR reconstruction gives the minimum L^2 reconstruction error and is thus preferable over the separate reconstruction (super-resolution for each HDR exposure and then HDR imaging to combine the HR frames). Moreover, separate registration for times which saturate a large part of the image can fail (*e.g.* the series of images from 4.8 (c) (f)).

4.5.1 Experiments

The following experiments were conducted in order to test the capacity of the proposed strategy to recover the dynamic and frequency information lost by a single image acquisition without the need of regularization hypothesis on the scene. To be as close as possible to our model we used the raw images from a high end commercial camera (Canon 7D) for real data experiments (second and third experiments)

Synthetic data In the experiment of Figures 4.6 and 4.7, we illustrate how taking the advocated number of images yields a better reconstruction than a total variation regularization with less images. We generate synthetic LR images, according to Model (4.2), from an HR irradiance map (extracted from a real scene). For TV regularization, we generate 2

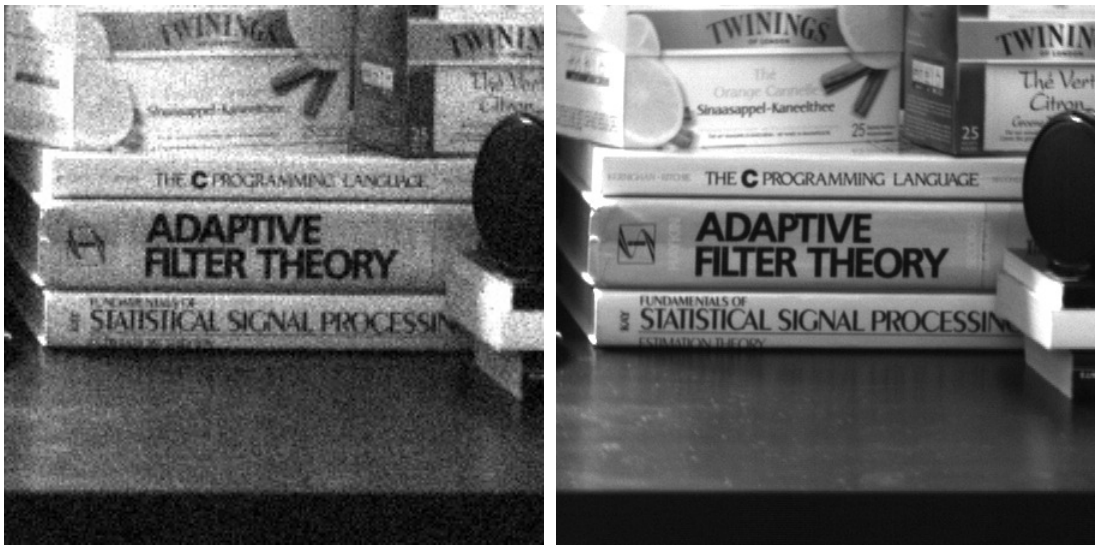
LR images at the three exposure times $\frac{1}{64}$, $\frac{1}{41}$ and $\frac{1}{32}$ seconds. We add a total variation regularization term to minimization (4.4) and choose the regularization parameter giving the best PSNR. For the unregularized case, we generate 20 images for each of these times, and perform minimization (4.4). The difference in reconstruction quality can be noticed both qualitatively, and in terms of PSNR. The unregularized version has a 6 dB PSNR improvement (43.48 dB for the regularized result and 49.92 dB for the proposed approach), and the content of the image is not altered (frequencies are not distorted). The TV regularized method altered edges, as for example in the small text areas.

High dynamic range planar surface We use a planar surface half illuminated by a strong source, thus generating two levels of irradiance. The scene is shown in Figure 4.5. Acquiring a plane matches the hypothesis of affine motion (small homography). In a more complex scene, segmenting the image in parts where the affine motion hypothesis is valid might be necessary.

The strategy introduced in section 4.5 is used first to find the 3 optimal exposure times (which are enough to capture the two different illumination levels of the scene) for the scene: $\frac{1}{100}$, $\frac{1}{80}$ and $\frac{1}{32}$ seconds. Then, 20 hand-held pictures are taken for each of the three selected exposures (total exposure time is approximately 1.1 second). Finally, the joint HDR-SR is performed solving (4.4) for all images.

Figure 4.8 shows example images for each exposure time in the irradiance domain (i.e. at the same scale) along with the corresponding saturation masks. We observe the increasing saturation for increasing exposure time. In Figure 4.9, the result of the HDR-SR reconstruction with factor $M = 2$ is displayed along with the result of a bi-cubic interpolation of the reference image for comparison purposes. All images are at the appropriate scale for a fair comparison. As expected, the HR image resulting from HDR-SR reconstruction is sharper and less noisy than the bi-cubic interpolation. In this practical case, the intrinsic frequency content of the scene is not very rich because the camera has a physical anti-aliasing filter that cuts much of the high frequencies. However, the recovery of higher frequency information (which does not rely on any regularity model of the scene) is visible, especially on the plot of frequency spectra.

Complex scene An HDR-SR experiment was conducted in the scene shown in Figure 4.10. The strategy introduced in section 4.5 is used to find 3 optimal exposure times for the scene: $\frac{1}{64}$, $\frac{1}{41}$ and $\frac{1}{32}$ seconds. Then 20 hand-held pictures are taken for each of the three selected exposures. The total exposure time is approximately 1.4 seconds. In order to apply (4.4), the HDR-SR reconstruction must be performed in sub-regions of the images verifying the affine motion hypothesis. The joint HDR-SR for each region is performed solving (4.4) from the corresponding regions of all input images. Figure 4.11 shows the HDR-SR reconstruction for a bright and a dark region of the scene. The recovery of high resolution content is particularly visible in the focused part of the image (see figures 4.11(d) and 4.12). In particular, the aliasing in the LR image (noticeable on the text) is greatly reduced in the reconstruction. The improvement is less noticeable in the bright region (see figure 4.11(b)) since that part of the image is out of focus.



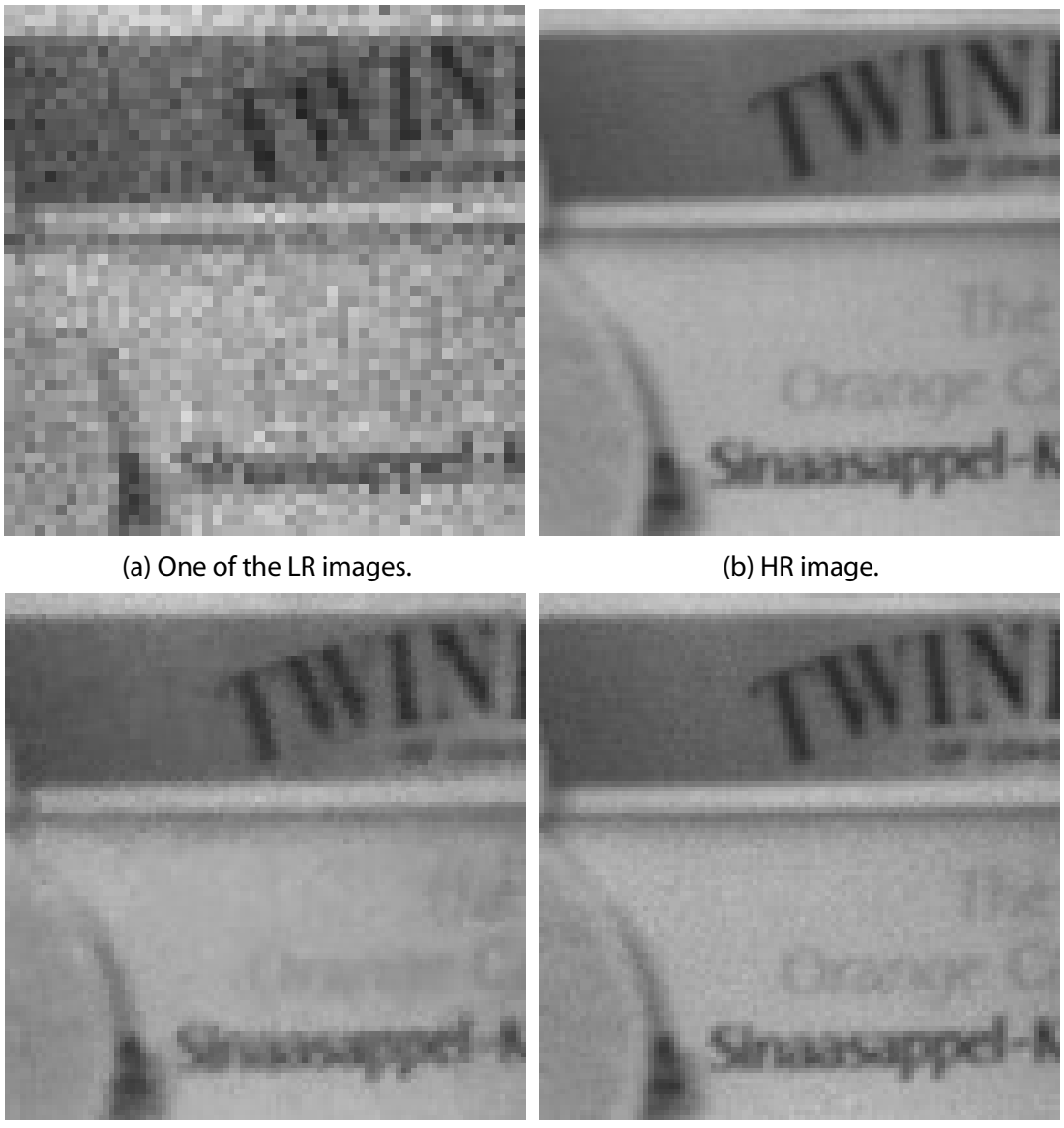
(a) One of the LR images.

(b) HR image.

(c) Optimal TV regularization with 6 images (PSNR=43.48 dB).

(d) SR without regularization and with 60 images (PSNR=49.92 dB).

Figure 4.6: Comparison between a regularized method with 6 images and the proposed non-regularized method with 60 images. See details in Figure 4.7.



(a) One of the LR images.

(b) HR image.

(c) Optimal TV regularization with 6 images (PSNR=43.48 dB).

(d) SR without regularization and with 60 images (PSNR=49.92 dB).

Figure 4.7: Details of Figure 4.7.

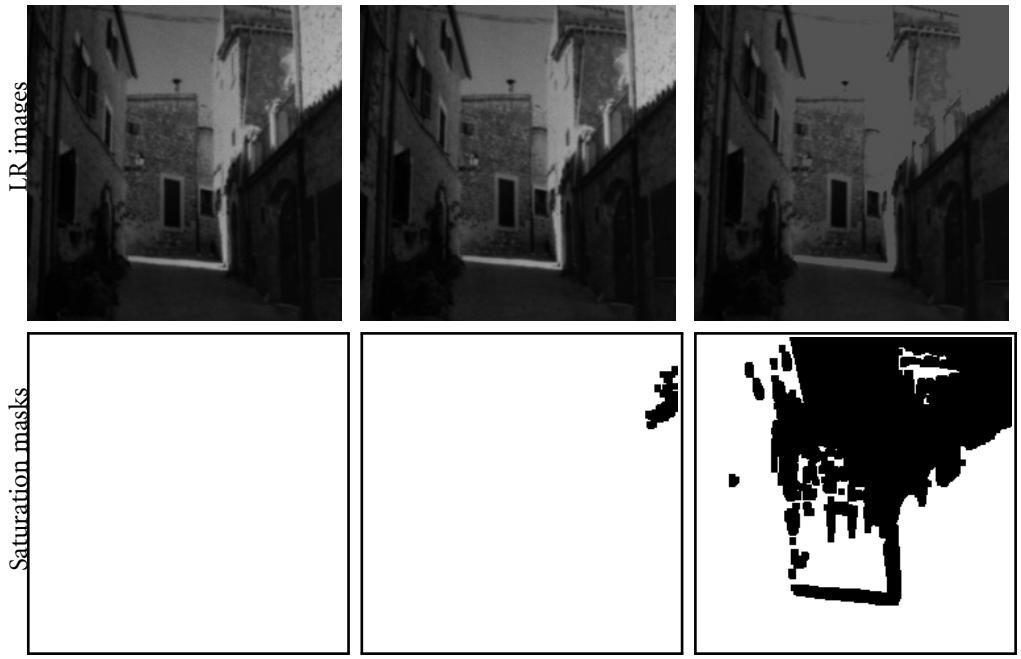


Figure 4.8: Acquired LR irradiance images with different exposure times with their respective saturation masks (saturated parts are in black). **First row (left to right):** reference image at $\frac{1}{100}$, image at $\frac{1}{80}$, image at $\frac{1}{30}$. **Second row:** saturation masks for the corresponding images above.

4.6 Conclusions

In this chapter, we exposed how the high dynamic range super-resolution (HDR-SR) problem can be set-up as a minimization problem including state of the art techniques from both sides of the problem. We showed that particular care is necessary when choosing acquisition parameters for HDR-SR. A balance between noise generation due to the conditioning of super-resolution and the noise intensity corresponding to the length of exposure times should be found. We also showed how exposure times for HDR-SR can be chosen. The main conclusion of the work presented in this chapter is the suggested strategy which ensures that, if the affine motion hypothesis holds and sufficiently long exposure time is available then all the information contained in the irradiance scene (both in amplitude and in frequency) is recovered.

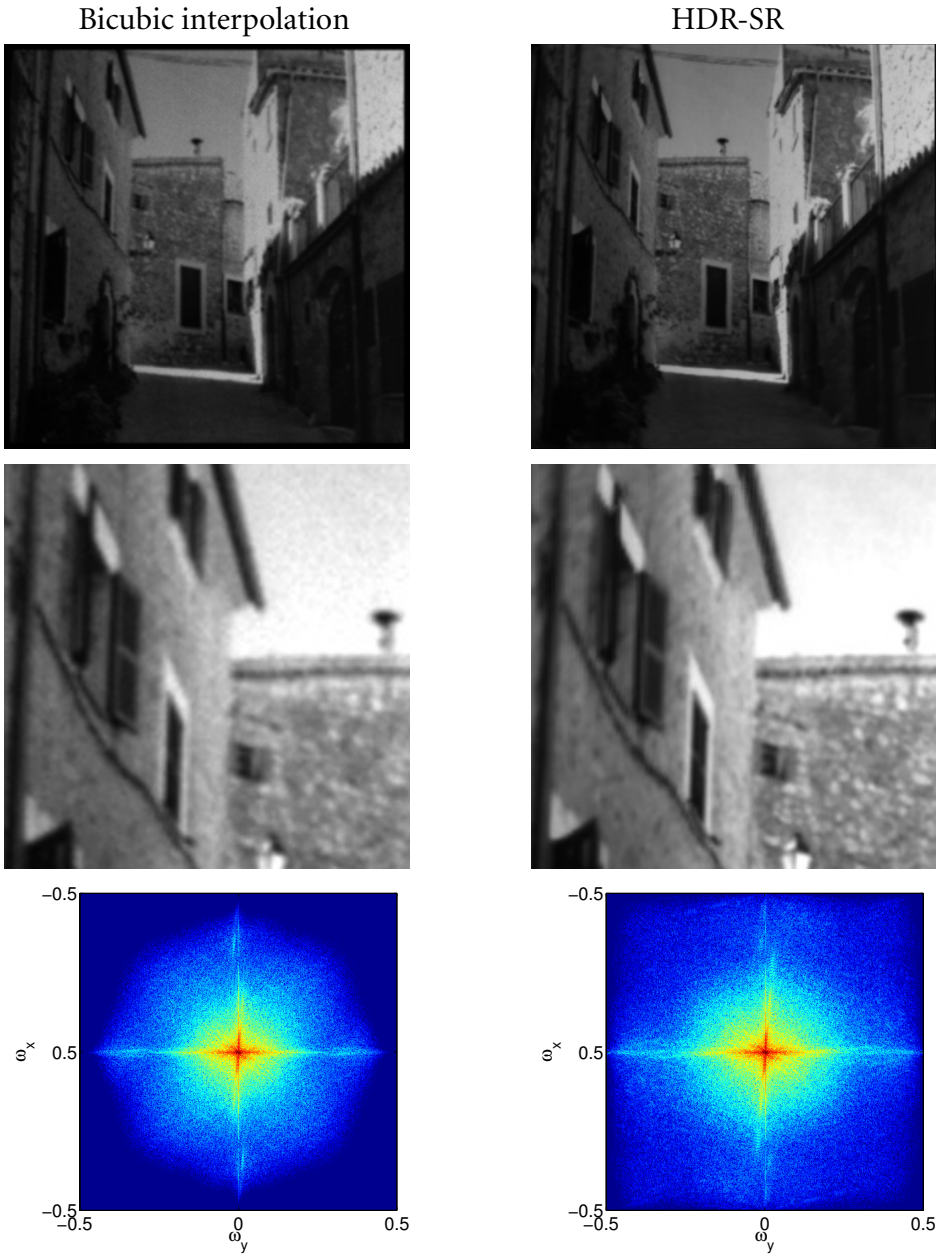
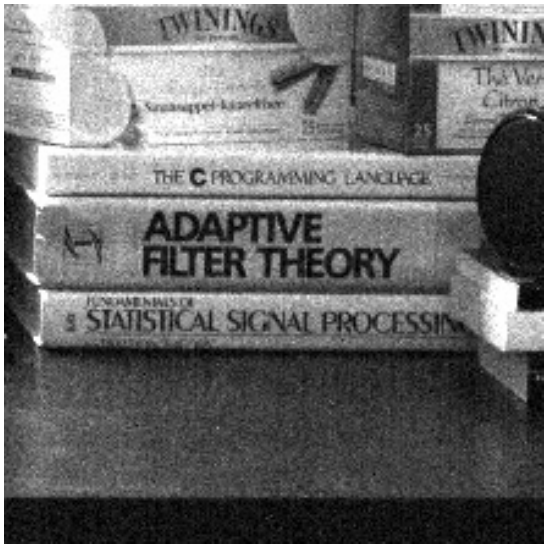


Figure 4.9: Result of HDR-SR with 60 images. **First column (top to bottom):** bicubic interpolation of the first LR image, detail of the bicubic interpolation, 2D frequency spectrum of the bicubic interpolation. **Second column (top to bottom):** result of the proposed HDR-SR approach, detail of the proposed HDR-SR result, 2D frequency spectrum of the proposed HDR-SR result.

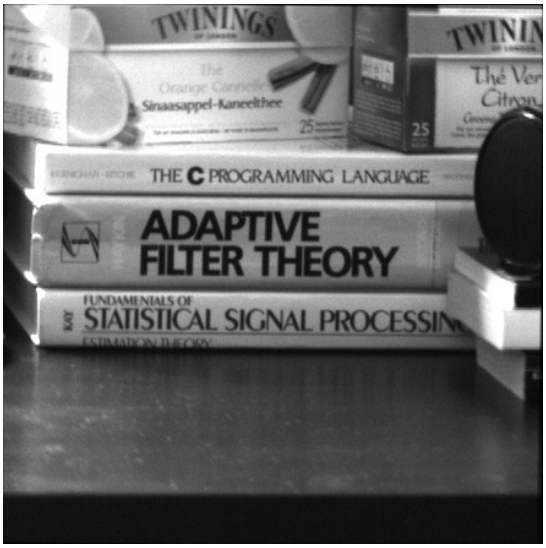
Figure 4.10: Scene at two different exposure times for our second experiment.

(a) LR image.

(b) Result of the proposed HDR-SR.



(c) LR image.



(d) Result of the proposed HDR-SR.

Figure 4.11: Result of the proposed HDR-SR method with 60 images.

(a) Detail of the result in Figure 4.11 (c). (b) Detail of the result in in Figure 4.11 (d).

Figure 4.12: Details of the result of the proposed HDR-SR method with 60 images shown in Figure 4.11.

5 HDR imaging for dynamic scenes and hand-held camera

As presented in Chapter 3, HDR images are usually generated by combining multiple photographs acquired with different exposure times. This approach, while effective, suffers from various drawbacks. The irradiance estimation is performed by combining, for each pixel, different exposure values at the same spatial position. This estimation scheme does not take advantage of the redundancy present in most images. Moreover, images must be perfectly aligned and objects must be in the exact same position in all frames in order to combine the different exposures without introducing ghosting artifacts. Figure 5.1 shows the result obtained when combining misaligned images. Ghosting artifacts appear all over the image giving a completely unusable result.

In this chapter, we present a new HDR imaging approach that simultaneously copes with these problems and exploits image redundancy to produce a denoised result. A reference image is chosen and a patch-based approach is used to find similar pixels that are then combined for the irradiance estimation. This patch-based approach permits to obtain a denoised result and is robust to image misalignments and object motion. Results show significant improvements in terms of noise reduction over previous HDR imaging techniques, while being robust to motion and changes between the exposures. The work presented in this chapter has been published in the proceedings of the International Conference on Computational Photography (ICCP) held at Boston, US, in April 2013 [Aguerrebere et al. 2013].

The chapter is organized as follows. We start in Section 5.1 with a summary of the previous work on HDR imaging for dynamic scenes. In Section 5.2 we start with a short reminder of the state-of-the-art in static HDR imaging techniques and then we describe the different steps of our non local HDR estimation approach. Experiments and comparisons with state-of-the-art approaches are presented in Section 5.3. Conclusions are presented in Section 5.4.

5.1 Previous work

There is a broad literature dedicated to the creation of HDR images for completely static scenes and static camera (c.f. Chapter 3). This problem is simplified but nevertheless

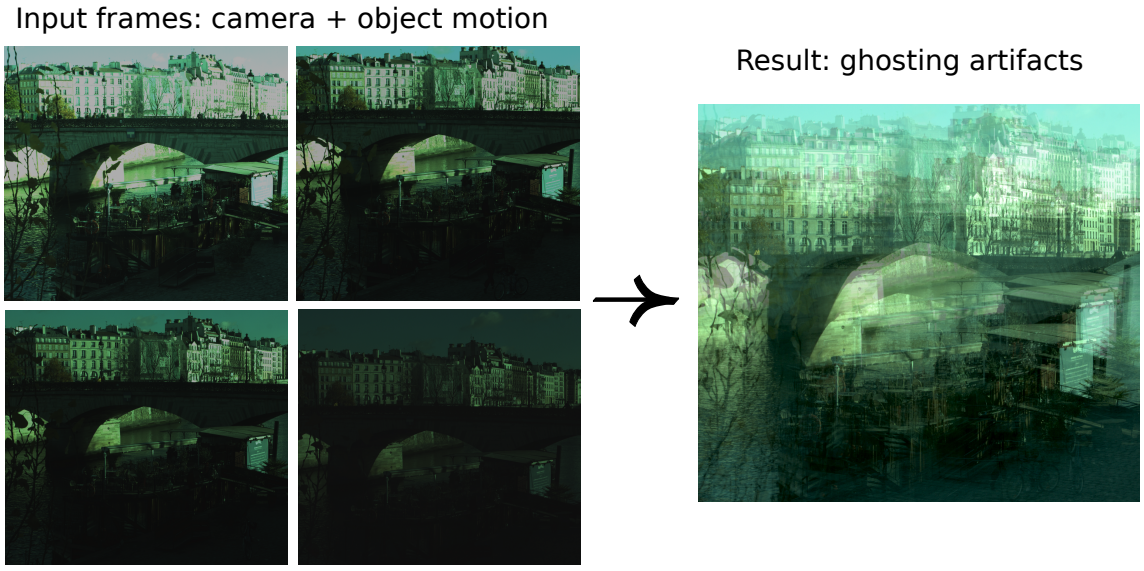


Figure 5.1: **Left:** Input misaligned images (these images are the result of demosaicking raw images, white balance has not been performed). **Right:** Result of the combination of the input frames without previous corregistration. Ghosting artifacts appear due to image misalignment.

particularly interesting. Knowing the physical model of the camera sensor and working with raw data, the most recent approaches [Kirk and Andersen 2006; Granados et al. 2010; Aguerrebere et al. 2014b] model the HDR creation as a statistical estimation problem, for which the maximum likelihood estimator is shown to be nearly optimal [Aguerrebere et al. 2014b](Chapter 3). While powerful from the theoretical point of view, these approaches greatly suffer in practice from small camera misalignments and cannot cope with moving objects, which makes them hardly reliable in non controlled environments.

Several approaches have been proposed in the literature to overcome these limitations. Most of them start by globally registering the frames in order to compensate for global camera motion. As a result, the background of the scene can be seen as static, while dynamic objects remain unregistered.

The first kind of methods makes use of optical flow techniques in order to compensate the motion across the different frames, generally after a global exposure compensation between the frames [Kang et al. 2003; Bogoni 2000; Zimmer et al. 2011]. Unfortunately, explicit motion estimation remains an ill-posed and ambiguous problem (especially in the presence of missing regions), and such methods are exposed to alignment errors.

A second variety of deghosting methods aims at reducing the influence of pixels of moving objects on the HDR estimation. These pixels can be detected in a first step and completely discarded for the estimation [Grosch 2006; Jacobs et al. 2008; Sidibé et al. 2009; Gallo et al. 2009]. Alternatively, the pixel weights in the irradiance estimation can be adapted in order to decrease the influence of suspicious pixels [Khan et al. 2006; Min et al. 2009; Pece and Kautz 2010; Heo et al. 2011]. For a recent and complete review on ghost detection and removal, see [Srikantha and Sidibé 2012].

Another limitation of HDR imaging is that images created from different exposures tend to be noisy. This comes from the fact that the HDR estimation is generally performed

on a per-pixel basis (possibly after registration). For a given pixel, the estimation only relies on the observed values at the same location for all the available exposures. Moreover, this number is reduced if a moving object occurs or if the pixel is over or underexposed in some frames. The recent work by Heo et al. [2011] proposes a first solution to this problem, making use of a bilateral filtering during the frame fusion step. However, such a local filtering cannot cope with large motions.

In this chapter, we propose to take advantage of the strong self-similarity property of images in order to solve the HDR estimation problem. More precisely, we rely on the redundancy of image patches to correctly estimate the irradiance at each pixel while avoiding the fastidious image registration and motion estimation (or detection) steps. In particular, the patch similarities are sufficiently robust to handle camera motion and the possibly large motion of objects. As an important by-product, the result of our non local HDR algorithm is much less noisy than the results of state-of-the-art HDR approaches. This denoising capacity is obviously true for the background scene, but also for objects in motion, whose different occurrences through the exposures are correctly retrieved thanks to the patch-based approach.

5.2 HDR imaging for dynamic scenes and hand-held camera

The methods presented in Chapter 3 assume that the images have been perfectly co-registered. These methods work under the hypothesis that the camera is fixed with respect to a static scene. As a consequence, their performance is highly affected by image misalignments and moving objects. Indeed, the HDR estimation problem becomes far more difficult in the case of dynamic scenes. A precise prior registration and some kind of motion estimation are seemingly essential to perform the frame combination if one wants to avoid blur and ghosting artifacts. Now, motion estimation is often subject to ambiguities and fine global registration becomes complicated for large camera motion.

In the following, we propose to solve the HDR estimation problem in the dynamic case by exploiting the inherent redundancy of images. Drawing on the recent works on multi-image denoising [Buades et al. 2008; Boracchi and Foi 2008], the idea is to combine information from pixels potentially sharing the same underlying irradiance value. In the process, we show that the HDR estimation of dynamic scenes does not require to apply any special or dedicated motion estimation method. Let us begin with a reminder of the HDR imaging technique that will then be needed in the proposed non-local approach.

5.2.1 HDR imaging for static scenes and camera

As presented in Chapter 3, several methods have been proposed to solve the HDR image generation problem following a pixel-wise approach in the case of static scenes. The basic idea, common to all of them, consists in combining T images acquired with exposure times $\tau_1, \ldots, \tau_T$. The irradiance C_p at each pixel p is then computed as a weighted average

of the irradiance estimated from each frame,

$$\hat{C}_p = \sum_{i=1}^{T} w_i^p \left(\frac{f^{-1}(\mathrm{z}_i^p)}{\tau_i} \right), \qquad \sum_{i=1}^{T} w_i^p = 1, \tag{5.1}$$

where f^{-1} is the inverse of the camera response function, z_i^p is the pixel value at position p for the frame acquired with exposure τ_i and w_i^p is the weight assigned to the i-th exposure for pixel p.

The most recent approaches [Granados et al. 2010; Aguerrebere et al. 2014b] propose to exploit raw camera values (in that case, f is linear before attaining the saturation threshold) and draw on a precise knowledge of the camera sensor noise to solve the irradiance estimation problem. Under this noise model, non saturated samples are seen as realizations of random variables Z_i^p distributed according to (c.f. Chapter 2)

$$\mathrm{Z}_i^p \sim \mathcal{N}(g a_p \tau_i C_p + \mu_R, g^2 a_p \tau_i C_p + \sigma_R^2), \tag{5.2}$$

where g is the camera gain, a_p models the photo response non uniformity (PRNU) factor, μ_R and σ_R^2 are the readout noise mean and variance.

Given that a closed formula for C_p under Model (2.2) cannot be found for the MLE, different numerical solutions have been proposed in the literature for its numerical estimation. The most efficient one is due to Granados et al. [2010], which propose an iterative algorithm: at each iteration, the irradiance $\hat{C}_p$ is computed as a weighted average of the irradiance estimations from each frame x_i^p with weights equal to the inverse of the variance of each estimation σ_{pi}^2,

$$\hat{C}_p = \frac{\sum_{i=1}^{T} \frac{\mathrm{x}_i^p}{\sigma_{pi}^2}}{\sum_{i=1}^{T} \frac{1}{\sigma_{pi}^2}}, \qquad \text{with} \qquad \mathrm{x}_i^p = \frac{\mathrm{z}_i^p - \mu_R}{g a_p \tau_i}, \qquad \sigma_{pi}^2 = \frac{g^2 a_p \tau_i \mathrm{x}_i^p + \sigma_R^2}{(g a_p \tau_i)^2}. \tag{5.3}$$

The weights are initialized directly from the input samples. Granados et al. also consider the dark currents which we neglect here for being substantially smaller than the other considered noise sources (c.f. Chapter 2).

The MLE approach by Granados et al. can be considered to be the state-of-the-art in pixel-wise HDR generation for static scenes. Moreover, we showed [Aguerrebere et al. 2014b] that this MLE approximation performs very close to the Cramér-Rao lower bound of the irradiance estimation problem. This result is proved for a pixel-wise estimation discarding the saturated samples and raw pixel values following Model (2.2). Also we showed [Aguerrebere et al. 2014b] that in most cases, the iterative approach of [Granados et al. 2010] does not require more than one iteration to yield good results, the estimated irradiance remaining almost unchanged after the first iteration.

5.2.2 The proposed non-local approach

The diagram presented in Figure 5.2 summarizes the proposed method. Given the T frames, we first choose a reference image (c.f. Section 5.2.2 for the reference frame selection) and estimate the irradiance on each pixel by combining the information of similar pixels present in all frames, i.e. the reference frame and the frames corresponding to

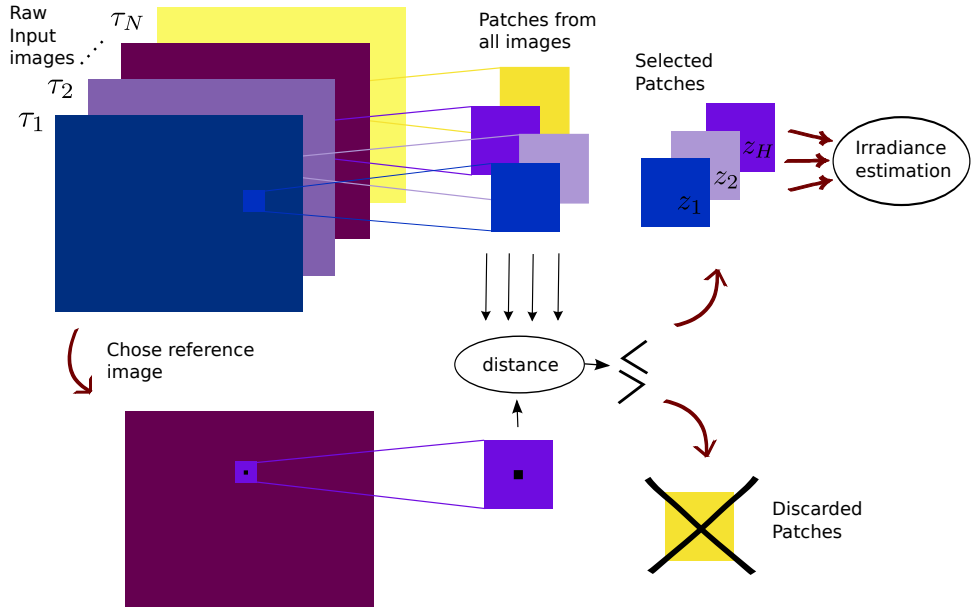


Figure 5.2: Diagram of the proposed method for HDR imaging with hand-held camera and moving objects.

the other exposure times. Similar pixels are found by comparing the neighborhood of the pixel in the reference image with those of pixels in all frames. Finding similar pixels through patch comparison makes the method robust to image misalignments, resulting from hand-held camera motion, and to moving objects without the need of image co-registration or motion detection. We propose to work with raw data in order to exploit the potential of the statistical noise model introduced in Chapter 2.

Finding similar pixels

The first step of the method is to determine which pixels share the same irradiance value. For this purpose we propose to use a variation of the patch comparison used by the NL-MEANS method [Buades et al. 2005a]. This variation is adapted to the noise Model (2.2). The original NL-MEANS algorithm aims to denoise images corrupted with Gaussian noise of constant variance. Under Model (2.2), the noise is Gaussian distributed but the variance depends on both the irradiance and the exposure time. Therefore, the noise variance is different for each pixel (different C_p) in each frame (different τ).

Patches must be normalized by the exposure time in order to be comparable. Thus the comparison is done in the irradiance domain. The proposed distance between a patch centered at pixel p in the reference frame and a patch centered at pixel q in another frame is given by

$$d(p,q) = \frac{1}{N}\sum_{j=1}^{N} \frac{(\mathrm{x}_{pj} - \mathrm{x}_{qj})^2}{2\sigma_{pj}^2}, \tag{5.4}$$

where x_{pj} (resp. x_{qj}) is the irradiance of the j-th pixel of the patch centered at pixel p

(resp. at q) in the reference frame (resp. in the other frame) and σ_{pj}^2 is the irradiance variance, given by

$$\mathrm{x}_{pj} = \frac{\mathrm{z}_{pj} - \mu_R}{g a_p \tau}, \qquad \sigma_{pj}^2 = \frac{g(\mathrm{z}_{pj} - \mu_R) + \sigma_R^2}{(g a_p \tau)^2}, \tag{5.5}$$

with z_{pj} the value of the j-th pixel of the patch centered at pixel p, τ its exposure time and N the number of non saturated pixels in the patch. Saturated samples are discarded and only those patches with at least a minimum percentage of non-saturated samples are considered. The camera parameters (g, a_p, μ_R and σ_R^2) needed to compute $d(p,q)$ are assumed to be known from a previous camera calibration step [Granados et al. 2010].

Distance thresholding It can be shown that if the two patches centered at p and q come from the same (un-noisy) underlying patch, the distance $d(p,q)$ is proportional to a random variable following a chi-squared distribution. This result is used to threshold the distance judiciously and compute binary weights for the patches.

If x_{pj} and x_{qj} come from the same underlying irradiance C_j, and we neglect the read-out noise variance σ_R^2 and the difference between the PRNU factors at different positions, from (5.13) we have

$$\sigma_{pj}^2 = \frac{C_j}{\tau_p} \qquad \text{and} \qquad \sigma_{qj}^2 = \frac{C_j}{\tau_q}. \tag{5.6}$$

with τ_p and τ_q the exposure times of the patches centered at positions p and q respectively, which may differ if the patches come from different frames. Hence,

$$d(p,q) = \frac{1}{N} \sum_{j=1}^{N} \frac{(\mathrm{x}_{pj} - \mathrm{x}_{qj})^2}{2\frac{C_j}{\tau_p}} \tag{5.7}$$

$$= \frac{(1 + \frac{\tau_p}{\tau_q})}{2N} \sum_{j=1}^{N} \frac{(\mathrm{x}_{pj} - \mathrm{x}_{qj})^2}{\frac{C_j}{\tau_p}(1 + \frac{\tau_p}{\tau_q})} \tag{5.8}$$

$$= \frac{(1 + \frac{\tau_p}{\tau_q})}{2N} \sum_{j=1}^{N} \frac{(\mathrm{x}_{pj} - \mathrm{x}_{qj})^2}{\sigma_{pj}^2 + \sigma_{qj}^2}, \tag{5.9}$$

where $\sum_{j=1}^{N} \frac{(\mathrm{x}_{pj} - \mathrm{x}_{qj})^2}{\sigma_{pj}^2 + \sigma_{qj}^2}$ follows a chi-squared distribution with mean N and variance $2N$. We can thus compute the mean and variance of the patch distance $d(p,q)$ as

$$\mu_{d_{(p,q)}} = \frac{(1 + \frac{\tau_p}{\tau_q})}{2N} N = \frac{(1 + \frac{\tau_p}{\tau_q})}{2}, \tag{5.10}$$

and

$$\sigma_{d_{(p,q)}}^2 = \left(\frac{(1 + \frac{\tau_p}{\tau_q})}{2N} \right)^2 2N = \frac{(1 + \frac{\tau_p}{\tau_q})^2}{2N}. \tag{5.11}$$

The distance thresholding is performed so as to keep patches within h standard deviations from the mean distance

$$d(p,q) \leq \mu_{d_{(p,q)}} + h\sigma_{d_{(p,q)}}. \tag{5.12}$$

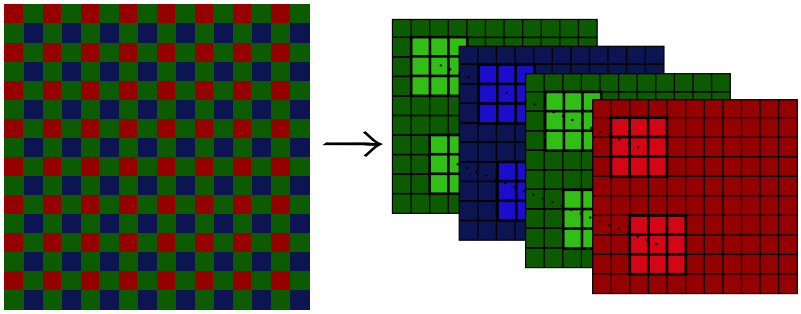


Figure 5.3: The Bayer pattern is decomposed into four sub-images in order to speed up the similar patch search.

If the distance between patches p and q is below the threshold, pixels p and q are assumed to follow the same distribution and thus q is used for the irradiance estimation of p (q is assigned weight 1). Otherwise q is discarded (q is assigned weight zero).

Irradiance estimation

The irradiance $\hat{C}_p$ at pixel p is computed as the first step of the MLE iterative approximation given by (5.3). The combined samples are the K pixels found to be similar to p according to the patch distance (5.4)

$$\hat{C}_p = \frac{\sum_{k=1}^{K} \frac{\mathrm{x}_p^k}{\sigma_{pk}^2}}{\sum_{k=1}^{K} \frac{1}{\sigma_{pk}^2}} \quad \text{with} \quad \mathrm{x}_p^k = \frac{\mathrm{z}_p^k - \mu_R}{g a_p \tau_k}, \qquad \sigma_{pk}^2 = \frac{g(\mathrm{z}_p^k - \mu_R) + \sigma_R^2}{(g a_p \tau_k)^2}, \tag{5.13}$$

where z_p^k is the k-th pixel found to be similar to pixel p and τ_k its exposure time.

Implementation details

The proposed algorithm works on raw data. Hence, the patch distance as well as the irradiance estimator must be computed combining pixels of the same color channel.

For each frame, the Bayer pattern is decomposed into four sub-images (red, blue and the two green channels separately considered). Patches are then considered as $n \times n \times 4$-dimensional, where n is the patch size and the third dimension represents the four color channels (see Figure 5.3). The distance between two $n \times n \times 4$-dimensional patches $d_T(p, q)$ is the mean distance among corresponding channels

$$d_T(p, q) = \frac{1}{4} \sum_{ch=1}^{4} d(p_{ch}, q_{ch}), \tag{5.14}$$

where $d(p_{ch}, q_{ch})$ is computed using (5.4). Note that this is equivalent to computing (5.4) between two patches of size $2n \times 2n$ in the original Bayer pattern. Then the two pixels are assumed to follow the same distribution if $d_T(p, q)$ is below a given threshold as specified in Section 5.2.2. In this way, a list of patches similar to p is created.

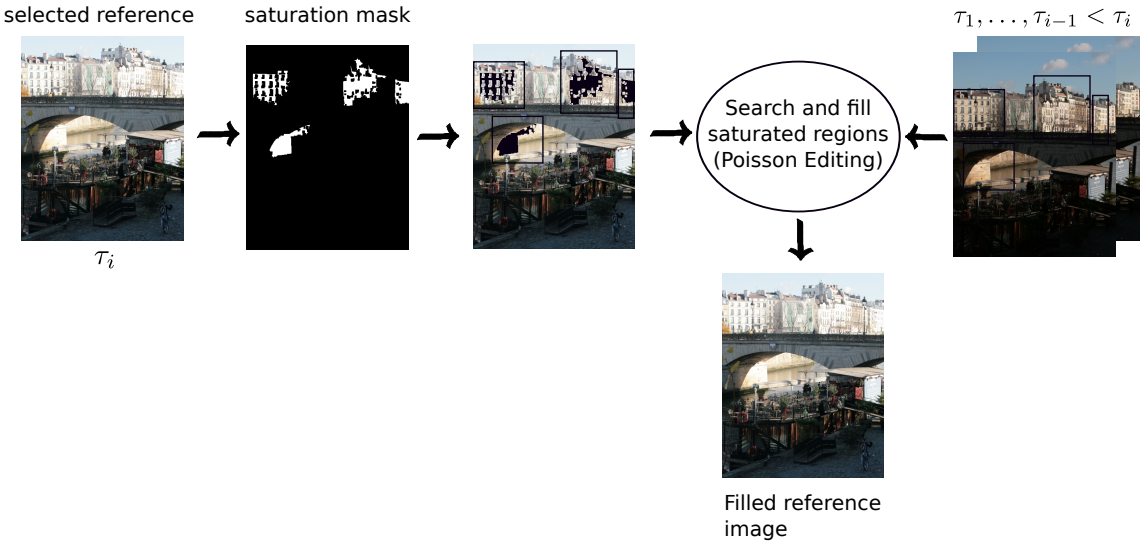


Figure 5.4: Filling of the reference image.

The central pixel of a patch corresponds to four different pixels in the Bayer pattern (one for each color channel). The irradiance estimator for each of these pixels is independently computed. For instance, the estimator of the red channel is computed combining the center pixels of the red components of selected patches. Note that the similar patches are found once and used to compute the four irradiance estimators. Another option is to do all processing on the Bayer image directly: center patches at each pixel, compute distances according to (5.4) and combine samples of the corresponding channels. The advantage of the proposed implementation is that the list of similar patches is computed once and used to estimate four irradiance values, thus making it four times faster.

Following [Buades et al. 2008] similar patches are sought on a limited search window. Since images are acquired with a hand held camera, a rough global translation estimation is first performed in order to compensate large global translations and allow a reasonable sized search window. The global euclidean distance between translated versions of each frame and the reference frame is computed in order to estimate the translation motion. Notice that this translation estimation is quite rough and only intends to correctly center the patch search window. In all our tests the search window is of size 11×11 pixels (which is equivalent to a window of 22×22 in the Bayer image).

Selection of the reference image

The selection of the reference image has a major impact on results. Indeed, the irradiance values of over or underexposed regions of the reference image cannot be recovered using the proposed method, since the patches lying on those regions do not contain reliable information. Thus these regions need to be identified and filled judiciously. We do not address this aspect of the problem in this work. We assume that it has been solved in a first stage by a well chosen inpainting technique, as for instance the ones described in [Wexler et al. 2007; Sen et al. 2012], which aim at filling these regions in the reference frame with

Figure 5.5: Tone mapped version of the HDR image used as ground-truth for the synthetic test presented in Section 5.3.1. From [Hasinoff et al.].

information taken from other frames. By doing so, we assume that an overexposed region of the reference frame will appear as not saturated in at least one of the shorter exposures. Conversely, we assume that underexposed regions can be retrieved from longer exposures. In this process, the reference image has to be chosen carefully, in order to limit the size of the regions to be inpainted. In practice, we choose this reference image as the shortest exposure containing no underexposed region. The HDR generation procedure described in Section 5.2 can then be directly applied using the filled reference image.

In this work, we make use of a classical inpainting technique, known as Poisson image editing [Pérez et al. 2003], in order to fill the overexposed regions in the reference image. The information is retrieved from other frames by comparing large patches surrounding the region to be filled with similar patches in the other frames. Figure 5.4 illustrates the filling process.

Another interesting possibility to complete the missing regions would be to rely on the work recently published by Sen et al. [2012], which proposes to generate an HDR image by choosing a reference frame and by filling the missing information through the minimization of a global energy. The unknown information is retrieved from the other frames keeping the coherence with the known regions of the reference image.

5.3 Results

5.3.1 Comparison to the case static scene / static camera

Releasing the hypothesis of aligned images is one of the key points that makes our method robust to camera motion and object motion. However, the price to pay may be quite high since that strong hypothesis carries a large amount of information implying that several samples per pixel can be reliably combined. In practice, that lost information is to be recovered by combining information from neighboring pixels.

In this section we present the results of an experiment conducted to evaluate the impact on performance of releasing this assumption on a perfectly static case. Four synthetic

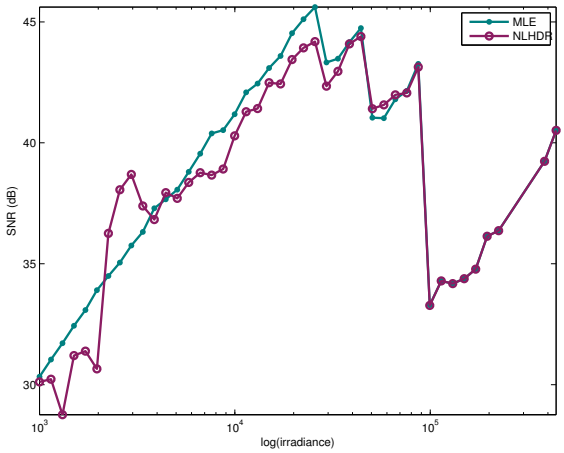


Figure 5.6: **Synthetic data.** SNR curves obtained using the classical MLE method [Granados et al. 2010] (green) and the proposed approach (violet). SNR values obtained with the non local approach are very close to those obtained with the classical method. Even though our approach releases the alignment hypothesis, it is close to the optimal result in a case where images are perfectly aligned.

images corresponding to four different exposures are generated according to Model (2.2) taking as irradiance ground-truth an HDR image. A tone mapped version of the ground-truth is shown in Figure 5.5. The camera parameters for the simulation are those of a Canon 400D set to ISO 200. This corresponds to the ideal case of the classical MLE approach: static camera and static scene. Figure 5.6 shows the SNR curves obtained using the classical MLE method [Granados et al. 2010] (green) and the proposed approach (violet). It can be verified that the SNR values obtained with the non local approach are very close to those obtained with the classical approach. Hence, releasing the alignment hypothesis does not have significant impact on the algorithm performance in a case where this hypothesis is perfectly satisfied.

This result was also verified using real data. Pictures of a static scene were acquired using a static camera remotely controlled from a laptop. The data set is composed of four images, corresponding to four exposure times acquired with a Canon 400D camera set to ISO 800. The first row of Figure 5.7 shows JPEG versions of the input image set[1]. Since our method treats raw data, a post-processing stage including demosaicking, white balance and tone mapping must be performed to display the results. To perform demosaicking we choose the technique by Hamilton and Adams [1997] because of its simplicity, it does not modify the known samples, and offers better performance than a simple bilinear interpolation. More evolved techniques were considered in order to avoid some artifacts present in Hamilton and Adams results. However, they were discarded since they also perform denoising of the samples which would interfere with our denoising results. For tone mapping we use the technique by Mantiuk et al. [2008].

Figure 5.7 shows the tone mapped version of the results obtained for the irradiance estimation of the scene. The second row shows the result obtained using the reference image

[1]All datasets and results are available at `http://perso.telecom-paristech.fr/~gousseau/hdr_denoising`

Figure 5.7: **Real data.** A static scene is acquired using a static camera. **First row:** Input images (JPEG version). **Second row:** Tone mapped reference image normalized by the exposure time. The saturated regions of the reference image were filled using the Poisson editing patch based approach described in Section 5.2.2. **Third row:** Tone mapped irradiance estimation by the MLE approach [Granados et al. 2010]. **Fourth row:** Tone mapped irradiance estimation of the proposed non local estimation approach. The result of the non local approach is less noisy than the one obtained using MLE. Thus releasing the alignment hypothesis does not have a significant impact on the algorithm performance. Recall that some demosaicking artifacts may appear due to the basic used technique. Please see electronic copy for better color and details reproduction.

Figure 5.8: **Real data.** Static scene except for one moving object (the postcard) acquired using a hand-held camera. **Left side. First row:** Input images (JPEG version). **Second row:** Tone mapped irradiance estimation using the proposed non local estimation approach. **Right side. First row:** Extracts of the moving object (the postcard). No ghosting artifacts appear. **Second / Third row:** Extracts of the normalized reference image (left), the method by Sen et al. [2012] (center) and the proposed non local approach (right). The irradiance estimation of the proposed approach is considerably less noisy than the one obtained using the reference frame only or the method by Sen et al.. Please see the electronic copy for better color and details reproduction.

normalized by the exposure time. The saturated regions of the reference image were filled using the Poisson editing patch based approach described in Section 5.2.2. We present this result for comparison purposes, since the normalized reference frame is the simplest HDR generation method we could use. The third row shows the results obtained using the MLE approach [Granados et al. 2010] and the fourth row shows the results obtained using the proposed non local estimation approach.

As expected, the noisier result is the one obtained with the normalized reference frame. On the other hand, the result for the non local approach is less noisy than the one obtained using MLE. The non local method gives a good quality result even if it does not make use of the alignment hypothesis. Sub-pixel motion (e.g. due to the camera's mirror) deviates samples from perfect alignment thus degrading the MLE performance with respect to the ideal case.

For this real example involving a static scene and a static camera, a ground-truth image could have been computed from several pictures so as to compare the SNR results of the different methods, in a way similar to [Granados et al. 2010; Hasinoff et al. 2010]. We declined to do so since, as shown in [Aguerrebere et al. 2014b](Chapter 3), a highly accurate knowledge of the camera model parameters is needed in order to obtain unbiased results, putting a strong question mark on the validity of such ground-truth.

5.3.2 Dynamic scene / Hand-held camera

As previously discussed, three main problems must be handled when generating an HDR image: camera motion (global motion), object motion and noise reduction of the irradi-

Figure 5.9: **Real data.** Dynamic scene (pedestrians in the bridge and people next to the boat) acquired using a hand-held camera. **Left:** Tone mapped irradiance estimation using the proposed non local approach. No ghosting artifacts appear. **Right first row:** Input images (JPEG version). **Right second row:** Extracts of the the normalized reference image. **Right third row:** Extracts of the results by Sen et al. [2012]. **Right fourth row:** Extracts of the results by the proposed non local approach. The result obtained by the proposed approach is far less noisy than the one by Sen et al.. Please see the electronic copy for better color and details reproduction.

ance estimation. Several methods addressing these problems can be found in literature. Nevertheless, to the best of our knowledge, none of them treats them simultaneously. There exists a wide variety of HDR methods that treat camera and object motion [Srikantha and Sidibé 2012]. In particular, the work recently presented by Sen et al. [2012], is shown to be among the state-of-the-art methods treating these issues.

On the other hand, the literature is less vast concerning methods for noise reduction of the irradiance estimation. The application of most classical denoising techniques is not straightforward since the irradiance image does not respect the usually considered hypothesis of additive white Gaussian noise with constant variance. Indeed, the noise distribution varies depending on the estimation method and the variance is seldom constant. Granados et al. [2010] propose to denoise the result of their irradiance estimation using a bilateral filter. Their irradiance estimation method allows to compute an estimate of the irradiance noise variance. Then the filtering parameter of the bilateral filter is set according to the estimated noise variance. Although Granados et al. apply some bilateral filtering after the MLE estimation, we chose not to include their results in the comparison because two critical hypothesis of their approach are not verified. Neither the camera nor the scene are fixed. Therefore, their results present severe ghosting problems on our test sets. In the following we present a comparative evaluation of methods performing HDR

imaging for dynamic scenes acquired with hand-held cameras. Next, we present some results comparing our simultaneous HDR - denoising technique against the denoising as a post-processing step after estimating the irradiance.

Evaluating simultaneous HDR imaging and denoising Here we present the results obtained by the proposed HDR image generation method on three sets of images. Each set is composed of pictures of a dynamic scene acquired with a hand-held camera. We compare our results to those obtained by Sen et al. [2012]. We also present the results obtained using the reference image normalized by the exposure time. The saturated regions of the reference image are filled using the Poisson editing patch based approach described in Section 5.2.2. We present this result for comparison purposes, since the normalized reference frame is the simplest HDR generation method we could use.

As already explained in Section 5.3.1, results are displayed using the demosaicking technique by Hamilton and Adams [1997] and the tone mapping technique by Mantiuk et al. [2008]. The first case is a static scene except for one moving object (the postcard). Images are acquired using a hand-held camera (Canon 400D set to ISO 800). The second row of Figure 5.8 shows a tone mapped version of the irradiance estimation obtained using our non local estimation approach. Three extracts of the results obtained for the moving object are shown in the first row (right side) of Figure 5.8. It can be verified that the result presents no ghosting artifacts. The second and third rows (right side) of Figure 5.8 present extracts of the results obtained with the normalized reference frame only (left), Sen et al. approach (center) and the proposed non local estimation method (right). The irradiance estimation of the proposed approach is far less noisy than the one obtained using the reference frame only or the method by Sen et al.. This example shows both, the effectiveness in noise reduction and the robustness to camera and object motion of the proposed approach.

Figures 5.9 and 5.10 present the results obtained for the two other test sets. These two cases present dynamic scenes where several objects are moving: pedestrians in the bridge and people next to the boat for the first example and pedestrians in the street and moving cars for the second example. Images are acquired using a hand-held camera (Canon 400D set to ISO 800 for the bridge scene and to ISO 400 for the street scene). The result for the proposed method is shown in Figures 5.9 and 5.10 (left). Extracts of the results obtained for the moving objects are presented in Figure 5.11 (left for the bridge scene and right for the street scene). It can be verified in both examples that no ghosting artifacts appear. The denoising capacity of the method is shown on various extracts. As in the previous example, the third and fourth rows of Figures 5.9 and 5.10 present extracts of the results obtained with the normalized reference frame only (left), Sen et al. approach (center) and the proposed non local estimation method (right).

It is interesting to remark that the non local approach is not likely to spread errors possibly introduced by the saturated region filling method. Recall that overexposed regions in the reference image must be filled before performing the non local estimation. If artifacts appear after hole filling, the created patches are not likely to appear elsewhere and thus wont have a significant effect on the estimation.

Figure 5.10: **Real data.** Dynamic scene (pedestrians in the street and moving cars) acquired using a hand-held camera. **Left:** Tone mapped irradiance estimation using the proposed non local approach. No ghosting artifacts appear. **Right first row:** Input images (JPEG version). **Right second row:** Extracts of the normalized reference image. **Right third row:** Extracts of the results by Sen et al. [2012]. **Right fourth row:** Extracts of the results by the proposed non local approach. The result obtained by the proposed approach is significantly less noisy than the one by Sen et al.. Please see the electronic copy for better color and details reproduction.

Figure 5.11: **Real data.** Extracts of the results obtained by the proposed non local estimation method for moving objects present on the scenes of Figures 5.9 (left) and Figure 5.10 (right). Notice that no ghosting artifacts appear in neither example.

On other ways to denoise irradiance maps Another possibility to denoise the irradiance map is to apply existing denoising techniques *after* estimating the irradiance. However, as previously mentioned, the application of most of these techniques is not straightforward since the irradiance image does not respect the usually made assumption of additive white Gaussian noise with constant variance. The noise distribution varies depending on the estimation method and the variance is seldom constant. This complicates the setting of parameters. Values chosen to correctly denoise bright regions will remove details in the dark regions while values correctly denoising dark regions will not denoise bright regions. To illustrate this behavior, the classical NL-MEANS algorithm[2] is applied to the normalized reference image and to the result obtained by Sen et al.. The NL-MEANS parameters are chosen so as to obtain a good visual compromise between removing noise and keeping details. Figure 5.12 shows extracts of the obtained results. The proposed approach manages to correctly denoise the bright regions (see the street panel) while still preserving details on the dark regions (see the tree branches). On the contrary, this is not the case for the two other examples. The denoising technique manages to remove part of the noise on the bright regions but at the cost of blurring the dark zones.

For the same reason, applying a classical denoising technique to each raw LDR image before the irradiance estimation is not a good option. Besides, information is lost when denoising each LDR image independently (e.g. details are lost, blurred edges) which might be kept in a multi-image denoising approach. Moreover, after denoising, the statistical model known for input samples is no longer valid. Thus the nearly optimal irradiance estimation obtained with the MLE estimator is no longer justified and an alternative optimal estimator should be found (which is not obvious given the new unknown model).

5.4 Conclusions

In this chapter we presented a new method for HDR image generation which copes simultaneously with three important problems: irradiance estimation noise, camera motion (hand-held camera) and multiple objects motion (dynamic scenes). Previous methods successfully handle these problems independently, but to the best of our knowledge none of them treats them all. The noise reduction capacity and robustness to camera and object motion of the proposed approach was experimentally verified in various real cases. The results show good denoising performance and no ghosting artifacts. As future work, an interesting research line is the improvement of the proposed algorithm including the results of the restoration method introduced in Chapter 6, which was not included at first due to the chronological order of development of both methods.

[2] Implementation from [Buades et al. 2011].

Figure 5.12: **Real data. Left:** Extracts from the NL-MEANS denoising of the normalized reference frame. **Center:** NL-MEANS denoising of the result by Sen et al.. **Right:** Proposed non local approach. The proposed approach manages to better denoise the bright regions (see the street panel) while better preserving details on dark regions (see the tree branches). On the contrary, this is not the case for the two other examples. The denoising technique manages to remove part of the noise on the bright regions but at the cost of blurring the dark zones. Only the green channel irradiance is displayed in order to avoid contrast changes introduced by the tone mapping techniques (needed to display HDR color images) and better visualize noise level differences.

6 A general restoration method with an application to single-image HDR

Image processing techniques aim at solving a wide variety of image restoration problems, from the classical denoising of additive Gaussian or impulsive noise and the retrieval of high frequency or high dynamic range information performed in super-resolution and HDR imaging, to the inpainting of missing pixels, either in large regions or for single missing pixels, the removal of blur, etc. These degradations are often combined in practice, as in raw digital images which present combined additive Gaussian noise and impulsive noise due to the acquisition process; the limited resolution and dynamic range also combined with noise as was treated in Chapter 4; or the missing pixels and noise present when using specialized acquisitions methods, as will be developed later in this chapter (see Figure 6.14, Section 6.4). In this work, we focus on the restoration of images simultaneously degraded by noise and random missing pixels.

Recent state-of-the-art methods make use of patch models and a Bayesian based reconstruction to restore degraded images. Some of them are devoted to the denoising problem [Lyu and Simoncelli 2009; Chatterjee and Milanfar 2012; Lebrun et al. 2013; Wang and Morel 2013], while others propose a more general framework for the solution of inverse problems [Zoran and Weiss 2011; Yu et al. 2012; Wang 2013a], including for instance inpainting, deblurring and super-resolution.

Patch models have been proven successful at image restoration and, in particular, Gaussian mixture models (GMM) have been proven accurate at representing natural image patches in the context of image restoration [Yu et al. 2012; Zoran and Weiss 2011]. Moreover, the Bayesian methods have been shown to be an adequate framework to exploit this patch model prior and to combine the observed and prior information in a fruitful manner. The work by Lebrun et al. [2013] presents a thorough and very interesting analysis of several recent restoration methods, revealing their a priori not obvious common roots and their relationship with the Bayesian approach.

Two methods attract our attention in the aim of tackling the combined problem of denoising and inpainting of random missing pixels, which as previously mentioned, finds application in several practical problems. First, the piecewise linear estimators (PLE) introduced by Yu et al. [2012], which constitutes a general framework for the resolution of inverse problems. Their approach makes use of a GMM to model image patches and a Bayesian method to restore the degraded patches. They obtain very promising results

in applications such as interpolation of missing pixels, deblurring, zooming and denoising. In particular, they show very good results in various cases of interpolation of random missing pixels with high masking rates (above 70% of missing pixels). Regarding the denoising problem, Lebrun et al. [2013] presented a non-local Bayesian based method (NLB), that also makes use of a Gaussian distribution prior on the image patches, and shows state-of-the-art results in several denoising examples with different noise levels. On the one hand, the denoising capacity of NLB is generally better than that of PLE [Wang 2013b], mostly due to the local estimation of the patch model as opposed to that of PLE which is done mainly globally in the image (actually regions of 128×128 are used). On the other hand, the NLB algorithm is a denoising algorithm and thus it is not made to interpolate missing pixels as PLE is. For these reasons, we propose a new Bayesian image restoration method, inspired by the combination of the PLE and NLB approaches, that tackles both problems simultaneously: image denoising and interpolation of random missing pixels, making use of a local Gaussian model for image patches.

We propose to model image patches according to a Gaussian prior, whose parameters will be estimated from local similar neighbors and a prior knowledge on the model. This way, we combine the denoising power of the local model estimation of NLB with the restoration power of PLE through the inclusion of a prior on the model. Then, the image patches are restored using the maximum a posteriori (MAP) estimator with the computed prior.

As will be made clear in the detailed explanation of the method in Section 6.3, unlike the PLE and NLB methods, the proposed approach deals with two different priors. One, also present in the PLE and NLB methods, is the prior on image patches. That is, a Gaussian model for image patches whose parameters, mean μ and covariance matrix $\mathbf{\Sigma}$, must be given or somehow estimated. The second, is a prior on μ and $\mathbf{\Sigma}$, used to compute a MAP estimate of these parameters. In Bayesian statistics, μ and $\mathbf{\Sigma}$ are known as hyperparameters, since they are the parameters of a prior distribution, while the prior on them is called an hyperprior. The use of an hyperprior allows the estimation of μ and $\mathbf{\Sigma}$ from similar samples even in the case of a large number of missing pixels in the patch.

The work presented here is an ongoing work and thus several points should be further analyzed and improved. Nevertheless, the results show that the proposed approach accurately reconstructs the missing pixels and denoise the known ones in various different conditions, ranging from low (20%) to high (70%) masking rates combined with different noise levels. This opens up the possibility of applying the proposed approach to the various restoration problems previously mentioned.

In particular, as will be presented in Section 6.4, the proposed method can be applied to the generation of HDR images when using a single-shot acquisition strategy with spatially varying pixel exposures (SVE). This acquisition technique, introduced by Nayar and Mitsunaga [2000], allows to capture HDR scenes with a single shot. This single-shot acquisition strategy overcomes the major drawbacks of the HDR multi-imaging techniques, i.e. the ghosting artifacts and global mis-alignment problems. However, it introduces new challenges to the HDR imaging problem. Unlike multi-image acquisition, where all pixels are correctly exposed in at least one of the images, in the SVE acquisition strategy the values of the saturated and under-exposed pixels are unknown and need to be

somehow restored. Moreover, noise is a bigger problem in SVE acquisition than it is in multi-imaging since there is considerably less information to help denoise the pixels (one single image). This is particularly troublesome in dark regions where most pixels are either under-exposed or close to become under-exposed and they have therefore a very low signal-to-noise ratio.

In Section 6.4 we introduce the SVE acquisition technique and explain how the presented Bayesian reconstruction method can be applied to restore the high dynamic range irradiance map from a single frame, simultaneously reconstructing the saturated and under-exposed pixels and denoising the correctly exposed ones.

The chapter is organized as follows. In Sections 6.1 and 6.2, we start with a brief summary of the two main methods, PLE and NLB, on which the new method is based. Section 6.3 is devoted to the presentation of the proposed approach. Section 6.4 presents the application of the proposed method to the generation of HDR images from a single shot acquired with spatially varying pixels exposures.

6.1 Piecewise linear estimators

Yu et al. [2012] introduced a general framework to solve image inverse problems using piecewise linear estimators (PLE). They propose to model image patches according to a Gaussian mixture model (GMM) and use the maximum a posteriori (MAP) estimator of the degraded patch to reconstruct the original. Let z_i be an observed degraded image patch of size $\sqrt{N} \times \sqrt{N}$, taken as a column vector of size N, considered to be drawn from the random variable Z_i,

$$\mathrm{Z}_i = \boldsymbol{D}_i \mathrm{C}_i + \mathrm{N}_i, \tag{6.1}$$

where $\boldsymbol{D}_i \in \mathbb{R}^{N \times N}$ is the degradation operator (e.g. random missing pixels), $\mathrm{C}_i \in \mathbb{R}^N$ is the original image patch we seek to estimate, $\mathrm{N}_i \in \mathbb{R}^N$ is the Gaussian distributed patch noise with zero mean and covariance matrix proportional to the identity matrix $\boldsymbol{\Sigma}_\mathrm{N} = \mathbf{I}\sigma^2$ and the index $i = 1, \ldots, I$ represents the patch position in the image.

The GMM describes image patches with K Gaussian distributions parametrized by their means μ_k and covariance matrices $\boldsymbol{\Sigma}_k$. Each patch C_i is assumed to be drawn independently from one of these Gaussians, whose probability density function is given by

$$p(\mathrm{C}) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Sigma}_k|^{1/2}} \exp\left(-\frac{1}{2}(\mathrm{C}-\mu_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathrm{C}-\mu_k)\right). \tag{6.2}$$

If the Gaussian parameters μ_k and $\boldsymbol{\Sigma}_k$ are assumed known, each patch $\tilde{\mathrm{C}}_i$ and its corresponding class $\tilde{k}_i$ can be computed as the MAP of the probability $p(\mathrm{C}_i|\mathrm{z}_i, \mu_k, \boldsymbol{\Sigma}_k)$

$$\begin{aligned}
(\tilde{\mathrm{C}}_i, \tilde{k}_i) &= \arg\max_{\mathrm{C},k} \ln p(\mathrm{C}|\mathrm{z}_i, \mu_k, \boldsymbol{\Sigma}_k) && (6.3)\\
&= \arg\max_{\mathrm{C},k} \left(\ln p(\mathrm{z}_i|\mathrm{C}, \mu_k, \boldsymbol{\Sigma}_k) + \ln p(\mathrm{C}, \mu_k, \boldsymbol{\Sigma}_k)\right) && (6.4)\\
&= \arg\min_{\mathrm{C},k} \left(\frac{\|\mathrm{z}_i - \boldsymbol{D}_i\mathrm{C}\|^2}{\sigma^2} + (\mathrm{C}-\mu_k)^T\boldsymbol{\Sigma}_k^{-1}(\mathrm{C}-\mu_k) + \ln|\boldsymbol{\Sigma}_k|\right). && (6.5)
\end{aligned}$$

The probability is first maximized over C and then over k. The solution $\tilde{\mathrm{C}}_i^k$ to the maximization over C is given by the Wiener filter

$$\tilde{\mathrm{C}}_i^k = \boldsymbol{\Sigma}_k \boldsymbol{D}_i^T (\boldsymbol{D}_i \boldsymbol{\Sigma}_k \boldsymbol{D}_i^T + \sigma^2 \mathbf{I})^{-1} (\mathrm{z}_i - \boldsymbol{D}_i \mu_k) + \mu_k. \tag{6.6}$$

The best model $\tilde{k}_i$ is then selected as the one minimizing (6.5) over k assuming $\mathrm{C} = \tilde{\mathrm{C}}_i^k$, that is

$$\tilde{k}_i = \arg\min_k \left(\frac{\|\mathrm{z}_i - \boldsymbol{D}_i \tilde{\mathrm{C}}_i^k\|^2}{\sigma^2} + (\tilde{\mathrm{C}}_i^k - \mu_k)^T \boldsymbol{\Sigma}_k^{-1} (\tilde{\mathrm{C}}_i^k - \mu_k) + \ln |\boldsymbol{\Sigma}_k| \right). \tag{6.7}$$

Because the Gaussian parameters μ_k and $\boldsymbol{\Sigma}_k$ are unknown, Yu et al. [2012] propose an iterative procedure to perform the reconstruction. The Gaussian parameters for the K classes are first initialized from a set of synthetic images covering a wide range of geometric configurations (see Section 6.1.1). At the so called *estimation step*, $\tilde{\mathrm{C}}_i$ and $\tilde{k}_i$ are computed according to equations (6.6) and (6.7) respectively. At the so called *model estimation step*, the classes parameters μ_k and $\boldsymbol{\Sigma}_k$ are updated computing the corresponding maximum likelihood estimators $\tilde{\mu}_k$ and $\tilde{\boldsymbol{\Sigma}}_k$ from the patches assigned to each class (the $\tilde{k}_i$ assigned at the previous step)

$$\tilde{\mu}_k = \frac{1}{|S_k|} \sum_{i \in S_k} \tilde{\mathrm{C}}_i, \quad \tilde{\boldsymbol{\Sigma}}_k = \frac{1}{|S_k|} \sum_{i \in S_k} (\tilde{\mathrm{C}}_i - \tilde{\mu}_k)(\tilde{\mathrm{C}}_i - \tilde{\mu}_k)^T, \tag{6.8}$$

with S_k the set of all patches assigned to class k and $|S_k|$ its cardinality. The covariance matrix $\tilde{\boldsymbol{\Sigma}}_k$ may be badly conditioned due for example to a small number of patches in the class. For this reason a regularization term $\varepsilon \mathbf{I}$ is added to ensure the correct inversion of the matrix [Yu et al. 2012]

$$\tilde{\boldsymbol{\Sigma}}_k = \tilde{\boldsymbol{\Sigma}}_k + \varepsilon \mathbf{I}. \tag{6.9}$$

The authors [Yu et al. 2012] analyzed the algorithm performance as a function of the iteration number and concluded empirically that 3 iterations are usually sufficient for convergence. The method is summarized in Algorithm 1.

In the experimental section of the work by Yu et al. [2012], the authors state that in all the presented examples, the proposed algorithm was applied to image regions of size 128×128. The input images are decomposed into regions of size 128×128, half-overlapped to avoid boundary effects, and the PLE algorithm is applied to each region. The estimates are averaged at the end to obtain the final reconstruction. The reason for this processing is that image content is usually more coherent locally than globally. Hence, it can be more accurately described with a limited number of classes if considered semi-locally than globally. This methodology is in line with the idea of performing a local processing as proposed by Lebrun et al. [2013] and the new approach here introduced. However, PLE remains a global approach if compared to the latter.

An interesting interpretation of this method, as described in [Yu et al. 2012], is that of the sparse modeling. The dictionary atoms are the eigenvectors of the K covariance matrices of the Gaussian models and a sparse representation of each patch is computed in this basis. This representation is the result of a non-linear step, that is the choice of the

Figure 6.1: **Left:** Example of synthetic contour black-and-white image used to initialize the PLE method. Patches that touch the contour at different positions are randomly sampled from it and used to compute the corresponding class covariance matrix. **Middle:** First 8 eigenvectors of the covariance matrix corresponding to the largest eigenvalues. **Right:** Typical eigenvalues used for initialization. Image from [Yu et al. 2012].

Gaussian model, and the linear projection of the patch into the chosen basis. The first non-linear step greatly reduces the degrees of freedom of the traditional sparse estimation methods which is otherwise K times larger.

6.1.1 PLE initialization

A good initialization is very important for algorithms, such as PLE, that try to solve a non-convex problem through an iterative approach. Yu et al. [2012] propose to initialize the algorithm using covariance matrices learned from synthetic images of edges with different orientations as well as the DCT basis to represent isotropic patterns. As they state, in dictionary learning, the most prominent atoms represent local edges which are useful at representing and restoring contours. Hence, this initialization helps to correctly restore corrupted patches even in quite extreme cases (e.g. more than 70% of missing pixels). Each class corresponds to one of $K-1$ orientations, uniformly sampled from directions zero to π. For a given orientation θ, a synthetic black-and-white image is generated and patches that touch the contour at different positions are randomly sampled from it. Figure 6.1 illustrates this process. A covariance matrix is then computed from the sampled patches. The first eigenvector of the covariance matrix, which is almost constant, is replaced by a constant vector. This allows a class of a given orientation to restore patches having different mean. Up to a certain gray level difference, dark or bright edges with the same orientation are correctly represented by the same class. A Gram-Schmidt orthogonalization is computed on the other eigenvectors to ensure the orthogonality of the basis. The eigenvalues of all bases are initialized with the same values, chosen to have a fast decay. At last, the DCT basis is added to represent isotropic image patterns, making a total of K classes. The mean of each class is initialized with zeros.

The authors claim that they have found in practice that 19 classes (i.e. 18 orientations, 10 degrees apart) give a correct reconstruction and are a good compromise between performance and complexity for a patch size of 8×8. All their experiments are performed with this fixed number of classes. The fact that the algorithm is applied in regions of size 128×128, and therefore localized, also explains why this a priori small number of classes can be suitable to describe all image patches.

Algorithm 1: Summary of the PLE algorithm.

```
Decompose the input image Z and D into overlapping patches.
Initialize the K Gaussian parameters μ_k and Σ_k.
for it = 1 to maxIts do
    for all patches do
        Compute C̃_i using (6.6).
        Compute k̃_i using (6.7) assuming C = C̃_i^k.
    end
    Build the reconstructed image from restored patches.
    Update μ_k and Σ_k using (6.8).
end
Build the final image from restored patches.
```

6.2 Non local Bayes

Non-local Bayes (NLB) [Lebrun et al. 2013] is an image denoising algorithm that restores image patches by computing an approximation of the MAP of the observed patches assuming a Gaussian prior. Unlike PLE, the Gaussian prior is here computed locally from the set of patches that are similar to the noisy patch and not from predefined classes.

Image patches are modeled according to (6.1) taking $\boldsymbol{D}_i$ as the identity matrix (all pixels are assumed known)

$$\mathrm{Z}_i = \mathrm{C}_i + \mathrm{N}_i, \tag{6.10}$$

and the MAP estimator $\tilde{\mathrm{C}}_i$ from the observed noisy patch z_i is computed as

$$\begin{aligned}
\tilde{\mathrm{C}}_i &= \arg\max_{\mathrm{C}} \ln p(\mathrm{C}|\mathrm{z}_i) && (6.11)\\
&= \arg\max_{\mathrm{C}} \ln p(\mathrm{z}_i|\mathrm{C}) + \ln(\mathrm{C}) && (6.12)\\
&= \arg\min_{\mathrm{C}} \left(\frac{\|\mathrm{z}_i - \mathrm{C}\|^2}{\sigma^2} + (\mathrm{C} - \mu_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathrm{C} - \mu_i) \right). && (6.13)
\end{aligned}$$

where μ_i and $\boldsymbol{\Sigma}_i$ are the Gaussian prior mean and covariance matrix respectively.

The NLB algorithm consists of two steps. In the first step, because μ_i and $\boldsymbol{\Sigma}_i$ are unknown, they are estimated by computing their maximum likelihood estimators $\tilde{\mu}_i$ and $\tilde{\boldsymbol{\Sigma}}_i$ from the set of patches found to be similar to the noisy patch z_i. If the selected patches are similar to those that would be obtained if compared to the original patch C_i, then the following approximation is valid

$$\tilde{\mu}_i \simeq \mu_i \qquad \text{and} \qquad \tilde{\boldsymbol{\Sigma}}_i \simeq \boldsymbol{\Sigma}_i + \sigma^2 \mathbf{I}. \tag{6.14}$$

Then, solving

$$\tilde{\mathrm{C}}_i = \arg\min_{\mathrm{C}} \left(\frac{\|\mathrm{z}_i - \mathrm{C}\|^2}{\sigma^2} + (\mathrm{C} - \tilde{\mu}_i)^T (\tilde{\boldsymbol{\Sigma}}_i - \sigma^2 \mathbf{I})^{-1} (\mathrm{C} - \tilde{\mu}_i) \right), \tag{6.15}$$

yields the patch reconstruction equation

$$\tilde{\mathrm{C}}_i = (\tilde{\mathbf{\Sigma}}_i - \sigma^2\mathbf{I})\tilde{\mathbf{\Sigma}}_i^{-1}(\mathrm{z}_i - \tilde{\mu}_i) + \tilde{\mu}_i. \tag{6.16}$$

Notice that (6.16) is very close to (6.6), except for the covariance matrix that is $(\tilde{\mathbf{\Sigma}}_i - \sigma^2\mathbf{I})$ instead of $\mathbf{\Sigma}_i$. This difference is due to the fact that, in the PLE algorithm the covariance matrix is computed from the restored patches while in the NLB algorithm it is computed from the noisy patches and it needs to be corrected for the extra noise term $\sigma^2\mathbf{I}$.

The selection of similar patches is crucial in the performance of the NLB algorithm since the quality of the Gaussian model parameters depends on it. The similar patches are found by computing the L^2 distance to all the patches in a neighborhood of it. A fixed number of patches is kept, which must be larger than the patch dimension to mitigate possible problems in the inversion of $\tilde{\mathbf{\Sigma}}_i$. Then the mean and covariance matrix are computed according to

$$\tilde{\mu}_i = \frac{1}{M}\sum_{j=1}^{M} \mathrm{q}_j^i, \tag{6.17}$$

and

$$\tilde{\mathbf{\Sigma}}_i = \frac{1}{M-1}\sum_{j=1}^{M} (\mathrm{q}_j^i - \tilde{\mu}_i)(\mathrm{q}_j^i - \tilde{\mu}_i)^T, \tag{6.18}$$

where q_j^i, $j = 1, \ldots, M$ is the set of M patches similar to z_i. Once the model parameters are estimated, all the similar patches are denoised simultaneously using (6.16) with the computed $\tilde{\mu}_i$ and $\tilde{\mathbf{\Sigma}}_i$. Finally, since each pixel belongs to several patches, its value is computed as the average of all the estimations of it coming from the denoised patches. This process, called aggregation, is usually used in patch-based reconstruction methods to rebuild the image from the restored patches.

The second step of the algorithm repeats the previous processing but takes the result of the first step as an oracle to find the similar patches and to improve the computation of the Gaussian parameters $\hat{\mu}_i$ and $\hat{\mathbf{\Sigma}}_i$

$$\hat{\mu}_i = \frac{1}{M'}\sum_{j=1}^{M'} \tilde{\mathrm{q}}_j^i, \tag{6.19}$$

and

$$\hat{\mathbf{\Sigma}}_i = \frac{1}{M'-1}\sum_{j=1}^{M'} (\tilde{\mathrm{q}}_j^i - \hat{\mu}_i)(\tilde{\mathrm{q}}_j^i - \hat{\mu}_i)^T, \tag{6.20}$$

where $\tilde{\mathrm{q}}_j^i$, $j = 1, \ldots, M'$ is the set of M' patches similar to $\tilde{\mathrm{C}}_i$.

Since the computation of $\hat{\mathbf{\Sigma}}_i$ is performed with the denoised patches, the approximation (6.14) is no longer needed and the reconstruction equation becomes

$$\tilde{\mathrm{C}}_i = \hat{\mu}_i + \hat{\mathbf{\Sigma}}_i(\hat{\mathbf{\Sigma}}_i + \sigma^2\mathbf{I})^{-1}(\mathrm{z}_i - \hat{\mu}_i). \tag{6.21}$$

Then, all the similar patches are denoised simultaneously using (6.21) and aggregation is performed to rebuild the image. The method is summarized in Algorithm 2.

Results from [Wang 2013b] show that NLB has a significantly better denoising power than PLE. The PLE algorithm, however, treats a wider range of inverse problems.

Algorithm 2: Summary of the NLB algorithm.

First Step:
Decompose the input image Z into overlapping patches.
Find patches $\mathrm{q}^i_{j=1,\dots,N}$ similar to current patch z_i according to the L^2 distance.
Compute the Gaussian parameters $\tilde{\mu}_i$ and $\tilde{\mathbf{\Sigma}}_i$ according to (6.17) and (6.18).
Denoise all similar patches $\mathrm{q}^i_{j=1,\dots,N}$ using (6.16) with the computed $\tilde{\mu}_i$ and $\tilde{\mathbf{\Sigma}}_i$.
Combine all restored patches to generate the reconstructed image $\tilde{\mathbf{C}}$.
Second Step:
Decompose the oracle $\tilde{\mathrm{C}}$ and the input image Z into overlapping patches.
Find patches $\tilde{\mathrm{q}}^i_{j=1,\dots,M}$ similar to current patch $\tilde{\mathrm{C}}_i$ according to the L^2 distance.
Compute the Gaussian parameters $\hat{\mu}_i$ and $\hat{\mathbf{\Sigma}}_i$ according to (6.19) and (6.20).
Denoise all similar patches $\mathrm{q}^i_{j=1,\dots,M}$ using (6.16) with the computed $\hat{\mu}_i$ and $\hat{\mathbf{\Sigma}}_i$.
Combine all restored patches to generate the reconstructed image.

6.3 Proposed approach: Hyperprior Non Local Bayes

The ability of PLE to interpolate missing pixels and the denoising power of NLB can be combined into a single algorithm that successfully treats images corrupted by additive Gaussian noise and random missing pixels. Even if the PLE algorithm is capable of performing both tasks, pixel interpolation and denoising, the superiority of NLB in terms of denoising motivates the goal to combine these approaches to improve the restoration performance of PLE. That is why we propose a new Bayesian image restoration method that tackles both problems: image denoising and interpolation of random missing pixels. The algorithm makes use of a local Gaussian model for image patches and an hyperprior on the model parameters. The use of an hyperprior allows the local estimation of the Gaussian model from similar samples even in the case of a large number of missing pixels in the patch. This way, we combine the denoising power of the local model of NLB with the restoration power of PLE. We refer hereafter to the proposed approach as Hyperprior Non Local Bayes (HPNLB).

6.3.1 Patch model

The considered patch model is a generalization of the Model (6.1). Recall that in (6.1), the noise term N_i is a vector of i.i.d. random variables following a Gaussian law with zero mean and constant variance σ^2, and thus independent of the signal C_i. We propose here to extend this model to consider the case of a noise term of independent random variables, but not necessarily equally distributed nor independent of C_i.

This more general framework opens up two possibilities: the consideration of noise with spatially variable variance yet independent of the signal C_i, and the more general case of noise dependent on C_i. The latter is particularly useful in the case of raw image data. As was introduced in Chapter 2, the noise present in raw image pixels is independent among pixels but its variance depends on the irradiance reaching the pixel, i.e. the noise

is dependent on the signal C_i.

The observed degraded patch $\mathrm{z}_i \in \mathbb{R}^N$ is assumed to be a realization of the random variable Z_i, following model

$$\mathrm{Z}_i = \boldsymbol{D}_i \mathrm{C}_i + \mathrm{N}_i, \tag{6.22}$$

where $\boldsymbol{D}_i \in \mathbb{R}^{N \times N}$ is the degradation operator, $\mathrm{C}_i \in \mathbb{R}^N$ is the original patch we seek to estimate, for which we assume a Gaussian prior $\mathcal{N}(\mu_i, \boldsymbol{\Sigma}_i)$, and $\mathrm{N}_i \in \mathbb{R}^N$ is a noise term, dependent on C_i, with its j-th entry given by

$$\mathrm{N}_i^j = f(\mathrm{C}_i^j)\varepsilon_i^j, \quad j = 1, \ldots, N, \tag{6.23}$$

where $(\varepsilon_i^j)_{j=1,\ldots,N}$ are i.i.d Gaussian random variables with zero mean and unit variance, ε_i^j is independent of $(\mathrm{C}_i^j)_{j=1,\ldots,N}$ and $f : \mathbb{R} \to \mathbb{R}$ is the function describing the relationship between the noise variance and the signal value at each pixel. The random variable N_i has a diagonal covariance matrix $\boldsymbol{\Sigma}_{\mathrm{N}_i}$. The matrices $\boldsymbol{D}_i$, $\boldsymbol{\Sigma}_{\mathrm{N}_i}$ and the function f are assumed to be known. For example, in the case of raw image data, the function f is given by

$$f(\mathrm{C}_i^j) = \sqrt{\frac{g^2 a_p \tau \boldsymbol{D}_i^j \mathrm{C}_i^j + \sigma_R^2}{(g a_p \tau)^2}}, \tag{6.24}$$

where g is the camera gain, a_p models the PRNU factor, τ is the exposure time, μ_R and σ_R^2 are the readout noise mean and variance (see Chapter 2, Equation (2.2), for a description of the raw image data model).

6.3.2 Patch restoration

As presented in Section 6.1, under Model (6.1) where the noise is independent of the signal and all considered vectors $(\mathrm{Z}_i, \mathrm{C}_i, \mathrm{N}_i)$ are Gaussian, the MAP estimator (6.6) matches the conditional mean estimator $\mathrm{E}[\mathrm{C}_i|\mathrm{Z}_i]$ and minimizes the Bayes risk $\mathrm{E}[(\tilde{\mathrm{C}}_i - \mathrm{C}_i)^2]$. This result cannot be directly applied to Model (6.22) because N_i is no longer independent of C_i nor Gaussian (due also to the dependence of N_i on the random variable $f(\mathrm{C}_i)$). Nevertheless, we show in the following that, among all *affine estimators* of C_i, the one that minimizes the Bayes risk under Model (6.22) is given by (6.6) (replacing the term $\sigma^2\mathbf{I}$ by $\boldsymbol{\Sigma}_{\mathrm{N}_i}$).

Proposition 7. *The affine estimator $\tilde{\mathrm{C}}_i$ that minimizes the Bayes risk $E[(\tilde{\mathrm{C}}_i - \mathrm{C}_i)^2]$ under Model* (6.22) *is given by*

$$\tilde{\mathrm{C}}_i = \boldsymbol{\Sigma}_i \boldsymbol{D}_i^T (\boldsymbol{D}_i \boldsymbol{\Sigma}_i \boldsymbol{D}_i^T + \boldsymbol{\Sigma}_{\mathrm{N}_i})^{-1} (\mathrm{Z}_i - \boldsymbol{D}_i \mu_i) + \mu_i. \tag{6.25}$$

Proof. Let us first consider the case $\mu_i = 0$. The estimator that minimizes the Bayes risk $\mathrm{E}[(\tilde{\mathrm{C}}_i - \mathrm{C}_i)^2]$ is equal to the conditional expected value $\mathrm{E}[\mathrm{C}_i|\mathrm{Z}_i]$. If we consider linear estimators only, we look for the matrix $\tilde{\mathbf{W}}$ that verifies

$$\tilde{\mathbf{W}} = \arg\min_{\mathbf{W}} \mathrm{E}[(\mathbf{W}\mathrm{Z}_i - \mathrm{C}_i)^2]. \tag{6.26}$$

Hence, $\tilde{\mathbf{W}}$ must verify

$$\mathrm{E}[(\tilde{\mathbf{W}}\mathrm{Z}_i - \mathrm{C}_i)\mathrm{Z}_i^T] = 0, \tag{6.27}$$

and we have

$$\tilde{\mathbf{W}} = \mathrm{E}[\mathrm{C}_i\mathrm{Z}_i^T](\mathrm{E}[\mathrm{Z}_i\mathrm{Z}_i^T])^{-1}. \tag{6.28}$$

From the patch Model (6.22), because ε_i^q is independent of C_i^p $\forall p, q = 1, \ldots, N$ and has zero mean we have

$$\mathrm{E}[\mathrm{C}_i^p\mathrm{N}_i^q] = \mathrm{E}[\mathrm{C}_i^p f(\mathrm{C}_i^q)\varepsilon_i^q] = \mathrm{E}[\mathrm{C}_i^p f(\mathrm{C}_i^q)]\mathrm{E}[\varepsilon_i^q] = 0. \tag{6.29}$$

Notice that since ε_i^q has zero mean we have

$$\mathrm{cov}(\mathrm{C}_i^p, \mathrm{N}_i^q) = \mathrm{E}[\mathrm{C}_i^p f(\mathrm{C}_i^q)\varepsilon_i^q] - \mathrm{E}[\mathrm{C}_i^p]\mathrm{E}[f(\mathrm{C}_i^q)\varepsilon_i^q] = 0, \tag{6.30}$$

thus (6.29) implies that C_i and N_i are not correlated. Hence, the element (p, q) of matrix $\mathrm{E}[\mathrm{C}_i\mathrm{Z}_i^T]$ is given by

$$\begin{aligned}
\mathrm{E}[\mathrm{C}_i\mathrm{Z}_i^T]_{p,q} &= \mathrm{E}[\mathrm{C}_i(\boldsymbol{D}_i\mathrm{C}_i + f(\mathrm{C}_i)\varepsilon_i)^T]_{p,q} && (6.31)\\
&= \mathrm{E}[\mathrm{C}_i^p(\boldsymbol{D}_i\mathrm{C}_i)_q + \mathrm{C}_i^p f(\mathrm{C}_i^q)\varepsilon_i^q] && (6.32)\\
&= \mathrm{E}[\mathrm{C}_i^p(\boldsymbol{D}_i\mathrm{C}_i)_q] + \mathrm{E}[\mathrm{C}_i^p f(\mathrm{C}_i^q)]\mathrm{E}[\varepsilon_i^q] && (6.33)\\
&= \mathrm{E}[\mathrm{C}_i^p(\boldsymbol{D}_i\mathrm{C}_i)_q] && (6.34)\\
&= (\boldsymbol{\Sigma}_k\boldsymbol{D}^T)_{p,q}. && (6.35)
\end{aligned}$$

Also, from the patch Model (6.22), the element (p, q) of matrix $\mathrm{E}[\mathrm{Z}_i\mathrm{Z}_i^T]$ is given by

$$\begin{aligned}
\mathrm{E}[\mathrm{Z}_i\mathrm{Z}_i^T]_{p,q} &= \mathrm{E}[(\boldsymbol{D}_i\mathrm{C}_i + f(\mathrm{C}_i)\varepsilon_i)(\boldsymbol{D}_i\mathrm{C}_i + f(\mathrm{C}_i)\varepsilon_i)^T]_{p,q} && (6.36)\\
&= \mathrm{E}[(\boldsymbol{D}_i\mathrm{C}_i)_p(\boldsymbol{D}_i\mathrm{C}_i)_q^T + (\boldsymbol{D}_i\mathrm{C}_i)_p(f(\mathrm{C}_i)\varepsilon_i)_q^T && (6.37)\\
&\quad + (f(\mathrm{C}_i)\varepsilon_i)_p(\boldsymbol{D}_i\mathrm{C}_i)_q + (f(\mathrm{C}_i)\varepsilon_i)_p(f(\mathrm{C}_i)\varepsilon_i)_q^T] && (6.38)\\
&= (\boldsymbol{D}_i\boldsymbol{\Sigma}_i\boldsymbol{D}_i^T)_{p,q} + (\boldsymbol{\Sigma}_{\mathrm{N}_i})_{p,q}. && (6.39)
\end{aligned}$$

Hence we have,

$$\tilde{\mathbf{W}} = \boldsymbol{\Sigma}_i\boldsymbol{D}_i^T(\boldsymbol{D}_i\boldsymbol{\Sigma}_i\boldsymbol{D}_i^T + \boldsymbol{\Sigma}_{\mathrm{N}_i})^{-1}. \tag{6.40}$$

In the general case, where $\mu_i \neq 0$, we can always consider the centered version of the patches $(\mathrm{z}_i - \boldsymbol{D}_i\mu_i)$ and apply the previous result to the affine case. Therefore, the estimator of C_i that minimizes the risk function $\mathrm{E}[(\tilde{\mathrm{C}} - \mathrm{C}_i)^2]$ among all affine estimators, under Model (6.22), is given by

$$\tilde{\mathrm{C}} = \boldsymbol{\Sigma}_i\boldsymbol{D}_i^T(\boldsymbol{D}_i\boldsymbol{\Sigma}_i\boldsymbol{D}_i^T + \boldsymbol{\Sigma}_{\mathrm{N}_i})^{-1}(\mathrm{z}_i - \boldsymbol{D}_i\mu_i) + \mu_i, \tag{6.41}$$

which matches (6.6) (replacing the term $\sigma^2\mathbf{I}$ by $\boldsymbol{\Sigma}_{\mathrm{N}_i}$). □

Even if this result is not a priori obvious due to the dependence between C_i and N_i, notice that it requires the non correlation of these variables (Equation (6.30)).

The same estimator is obtained if we compute the maximum of the posterior probability ignoring the dependence of $\boldsymbol{\Sigma}_{\mathrm{N}_i}$ on C_i. Then an iterative procedure is needed to update C_i and $\boldsymbol{\Sigma}_{\mathrm{N}_i}$ at each step. Without ignoring the dependence of $\boldsymbol{\Sigma}_{\mathrm{N}_i}$ on C_i, the minimization problem does not have an explicit solution and an approximate solution should be iteratively computed.

6.3.3 Estimation of the Gaussian parameters μ and Σ

Since μ_i and Σ_i are unknown in practice, a method must be found to estimate them. Both PLE and NLB propose to estimate these parameters computing their maximum likelihood estimators. In the case of PLE, the MLE is computed from all the patches that belong to the given class. In the case of NLB, the parameters are computed from the set of similar patches chosen among the neighbors of the current patch.

In this work, instead of using the MLE, we propose to compute the MAP of the class parameters μ_i and Σ_i using a set of similar patches and a prior knowledge on them. In Bayesian statistics, μ_i and Σ_i are known as hyperparameters, since they are the parameters of a prior distribution, while the prior on them is called an hyperprior. Using similar patches brings the advantages of the local estimation of NLB, while including the hyperprior allows to estimate the parameters in the case of missing pixels, which is possible with PLE but not with the NLB method.

In order to estimate μ and Σ we propose to compute their MAP estimators using a set of M patches found to be similar to the current patch (according to the L^2 distance) and an hyperprior on μ and Σ. We will come back to the precise choice of these patches at the end of the paragraph. For calculations simplicity we work with the precision matrix $\Lambda = \Sigma^{-1}$ instead of the covariance matrix Σ. As it is usual when considering hyperpriors, we rely on a conjugate distribution. In our case, that boils down to assuming a Normal-Wishart[1] prior for the couple (μ, Λ)

$$p(\mu, \Lambda) = \mathcal{N}(\mu|\mu_0, (\kappa\Lambda)^{-1})\mathcal{W}(\Lambda|(\nu\Sigma_0)^{-1}, \nu) \tag{6.42}$$

$$\propto \quad |\Lambda|^{1/2} \exp\left(-\frac{\kappa}{2}(\mu-\mu_0)\Lambda(\mu-\mu_0)^T\right) |\Lambda|^{(\nu-d-1)/2} \exp\left(-\frac{1}{2}\mathrm{tr}(\nu\Sigma_0\Lambda)\right), \tag{6.43}$$

where μ_0 is a prior on μ, Σ_0 is a prior on Σ, and $\kappa, \nu \in \mathbb{R}$, $\kappa > 0$, $\nu > d-1$. The matrix Σ_0 needs to be multiplied by ν in (6.43) since the expected value of Λ under the Normal-Wishart distribution is ν times its prior $(\nu\Sigma_0)^{-1}$.

The likelihood of the M patches $(\mathrm{z}_1, \ldots, \mathrm{z}_M)$ is given by

$$p(\mathrm{z}_1, \ldots, \mathrm{z}_M|\mu, \Lambda) = (2\pi)^{-d/2} \prod_{j=1}^{M} |\Lambda_j^*|^{1/2} \exp\left(-\frac{1}{2}\sum_{j=1}^{M}(\mathrm{z}_j - \boldsymbol{D}_j\mu)\Lambda_j^*(\mathrm{z}_j - \boldsymbol{D}_j\mu)^T\right). \tag{6.44}$$

with $\Lambda_j^* = (\boldsymbol{D}_j\Lambda^{-1}\boldsymbol{D}_j^T + \Sigma_w)^{-1}$. Then the MAP estimates $\hat{\mu}$ and $\hat{\Lambda}$ are found by maxi-

[1]The Normal-Wishart distribution is the conjugate prior of a multivariate normal distribution with unknown mean and covariance matrix. $\mathcal{W}$ denotes the Wishart distribution.

mizing the posterior probability

$$(\hat{\mu}, \hat{\mathbf{\Lambda}}) = \arg\max_{\mu, \mathbf{\Lambda}} p(\mu, \mathbf{\Lambda} | \mathrm{z}_1, \ldots, \mathrm{z}_M) \tag{6.45}$$

$$= \arg\max_{\mu, \mathbf{\Lambda}} p(\mathrm{z}_1, \ldots, \mathrm{z}_M | \mu, \mathbf{\Lambda}) p(\mu, \mathbf{\Lambda}) \tag{6.46}$$

$$= \arg\max_{\mu, \mathbf{\Lambda}} \prod_{j=1}^{M} |\mathbf{\Lambda}_j^*|^{1/2} \exp\left(-\frac{1}{2} \sum_{j=1}^{M} (\mathrm{z}_j - \boldsymbol{D}_j \mu) \mathbf{\Lambda}_j^* (\mathrm{z}_j - \boldsymbol{D}_j \mu)^T \right) \tag{6.47}$$

$$|\mathbf{\Lambda}|^{1/2} \exp\left(-\frac{\kappa}{2} (\mu - \mu_0) \mathbf{\Lambda} (\mu - \mu_0)^T \right) |\mathbf{\Lambda}|^{(\nu - d - 1)/2} \exp\left(-\frac{1}{2} \mathrm{tr}(\nu \mathbf{\Sigma}_0 \mathbf{\Lambda}) \right). \tag{6.48}$$

Computing the partial derivatives of $p(\mu, \mathbf{\Lambda} | \mathrm{z}_1, \ldots, \mathrm{z}_N)$ with respect to μ and $\mathbf{\Sigma}$ and equating to zero we find (c.f. Appendix B)

$$\hat{\mu} = \left(\kappa \mathbf{I} + \sum_{j=1}^{M} \mathbf{\Lambda}^{-1} \boldsymbol{D}_j^T \mathbf{\Lambda}_j^* \boldsymbol{D}_j \right)^{-1} \left(\sum_{j=1}^{M} \mathbf{\Lambda}^{-1} \boldsymbol{D}_j \mathbf{\Lambda}_j^* \mathrm{z}_j + \kappa \mu_0 \right) \tag{6.49}$$

and

$$\hat{\mathbf{\Lambda}}^{-1} = \boldsymbol{H_1} \boldsymbol{H_2}, \tag{6.50}$$

with

$$\boldsymbol{H_1} = \left((\nu - d) \mathbf{I} + \sum_{j=1}^{M} \mathbf{\Lambda}^{-1} \boldsymbol{D}_j^T \mathbf{\Lambda}_j^* \boldsymbol{D}_j \right)^{-1} \tag{6.51}$$

and

$$\boldsymbol{H_2} = \left(\sum_{j=1}^{M} (\mathbf{\Lambda}^{-1} \boldsymbol{D}_j \mathbf{\Lambda}_j^* (\mathrm{z}_j - \boldsymbol{D}_j \mu)) (\mathbf{\Lambda}^{-1} \boldsymbol{D}_j \mathbf{\Lambda}_j^* (\mathrm{z}_j - \boldsymbol{D}_j \mu))^T \right. \tag{6.52}$$

$$\left. + \kappa (\mu - \mu_0)(\mu - \mu_0)^T + \nu \mathbf{\Sigma}_0 \right). \tag{6.53}$$

From (6.49) we find that the MAP estimator of μ is a weighted average of two terms: the mean estimated from the similar patches $(\mathrm{z}_1, \ldots, \mathrm{z}_N)$ after projecting them into the "non-degraded" space

$$\sum_{j=1}^{M} (\mathbf{\Lambda}^{-1} \boldsymbol{D}_j \mathbf{\Lambda}_j^*) \mathrm{z}_j, \tag{6.54}$$

and the prior μ_0. The parameter κ controls the confidence level on the prior. Notice that the denominator term

$$\sum_{j=1}^{M} (\mathbf{\Lambda}^{-1} \boldsymbol{D}_j \mathbf{\Lambda}_j^*) \boldsymbol{D}_j \tag{6.55}$$

is the sum of the projected masks, giving the weights of the projected patches used in (6.54) to compute the patch mean.

With the same idea, we observe that the MAP estimator for $\mathbf{\Lambda}$ is a combination of the covariance matrix estimated from the projected patches

$$\sum_{j=1}^{M}(\mathbf{\Lambda}^{-1}\boldsymbol{D}_j\mathbf{\Lambda}_j^*(\mathrm{z}_j - \boldsymbol{D}_j\mu))(\mathbf{\Lambda}^{-1}\boldsymbol{D}_j\mathbf{\Lambda}_j^*(\mathrm{z}_j - \boldsymbol{D}_j\mu))^T, \tag{6.56}$$

the covariance imposed by μ (since μ and $\mathbf{\Lambda}$ are not independent),

$$(\mu - \mu_0)(\mu - \mu_0)^T \tag{6.57}$$

and the prior on $\mathbf{\Lambda}$ (or equivalently on $\mathbf{\Sigma}^{-1}$),

$$\mathbf{\Sigma}_0^{-1}. \tag{6.58}$$

Since (6.49) depends on $\mathbf{\Lambda}$ and (6.50) depends on μ and $\mathbf{\Lambda}$ those are not closed forms for the estimators. Hence, we propose to compute them using an iterative approach, defining an initial value for $\mathbf{\Lambda}$. As in the case of PLE, we have also observed that in practice 3 to 4 iterations is usually enough. Algorithm 3 summarizes the proposed method for the computation of $\hat{\mu}$ and $\hat{\mathbf{\Lambda}}$.

Algorithm 3: Computation of $\hat{\mu}$ and $\hat{\mathbf{\Lambda}}$.

1 Set parameters: $\mu_0, \mathbf{\Sigma}_0, \kappa, \nu$.
2 Initialize $\mathbf{\Lambda} = \mathbf{\Sigma}_0^{-1}$.
3 Compute $\hat{\mu}$ according to (6.49).
4 Set $\mu = \hat{\mu}$.
5 **for** *it = 1 to maxIts* **do**
6 Compute $\hat{\mathbf{\Lambda}}$ according to (6.50).
7 Set $\mathbf{\Lambda} = \hat{\mathbf{\Lambda}}$.
8 **end**

Search for similar patches The set of similar patches $(\mathrm{z}_1, \ldots, \mathrm{z}_M)$ are those with L^2 distance to the current patch below a given threshold. The threshold is set to a tolerance parameter β times the distance to the closest neighbor.

Since the patch comparison is performed in an oracle image (see details in Section 6.3.5), all pixels are assumed known. However, it may be useful to assign different confidence levels to the known pixels ($\boldsymbol{D}_p^j = 1$) and to those originally missing and then interpolated ($\boldsymbol{D}_p^j = 0$). For all the experimental results presented in Section 6.3.7, the distance between patches p and q is computed according to

$$d(p,q) = \frac{\sum_{j=1}^{N}(\mathrm{z}_p^j - \mathrm{z}_q^j)^2\omega_{p,q}^j}{\sum_{j=1}^{N}\omega_{p,q}^j}, \tag{6.59}$$

with

$$\omega_{p,q}^j = \begin{cases} 1 & \text{if } \boldsymbol{D}_p^j = \boldsymbol{D}_q^j = 1, \\ 0.01 & \text{otherwise.} \end{cases} \tag{6.60}$$

With this formulation, we are highly prioritizing the known pixels compared to the unknown ones. Variations of these weights should be tested. Moreover, this is an arbitrary way to define the patch distance, which should be revised in future work. In particular, it may be useful to consider a patch distance normalized by the pixel's variance, as performed in Chapter 5 (Equation (5.4)), in order to account for the spatially variable pixel noise.

6.3.4 Analysis of the parameters

Four parameters must be defined in order to compute μ and $\mathbf{\Sigma}$ using (6.49) and (6.50). That is the four parameters of the Normal-Wishart distribution: κ, ν, the prior mean μ_0 and the prior covariance matrix $\mathbf{\Sigma}_0$.

Setting of μ_0 and $\mathbf{\Sigma}_0$ Assuming an *oracle* is available (e.g., in an iterative approach, the oracle is the result of the previous iteration), μ_0 and $\mathbf{\Sigma}_0$ can be computed using the corresponding MLE estimators from the set of similar patches (Formulas (6.19) and (6.20)). Notice that this is the same approach taken by NLB, which is possible here despite the presence of unknown samples, since they are reconstructed in a previous iteration of the algorithm using (6.25).

It is clear though that the initialization of the algorithm plays a key role, since a first estimate of the unknown samples is needed in order to compute the corresponding μ_0 and $\mathbf{\Sigma}_0$. For that purpose, we make use of the initialization scheme proposed by Yu et al. [2012] for the PLE algorithm (c.f. Section 6.1).

The initialization of the proposed algorithm goes as follows. A fixed number of classes K is defined and the corresponding covariance matrices are computed from $(K-1)$ synthetic images of edges of different orientations plus the DCT to represent isotropic patterns as explained in Section 6.1. The means of the classes are set to zero. Each image patch is projected in each class using (6.6) and its best suited class is chosen using (6.7). A first oracle is thus created aggregating the estimations of all patches corresponding to the chosen class. This oracle is then used to initialize the proposed algorithm and compute the MLE estimates of μ_0 and $\mathbf{\Sigma}_0$.

Setting of κ and ν The estimation of μ is a combination of the mean estimated from the similar patches after projection into the *non-degraded* space through (6.25) and the prior mean μ_0. The parameter κ is related to the degree of confidence we have on the prior μ_0. Hence, its value should be a trade-off between the confidence we have in the prior accuracy and the confidence we have in the information provided by the similar samples. The latter is related to the number of similar neighbors and the number of known pixels in the current patch. A higher κ is needed if few similar patches are available or a large part of the patch is unknown.

Similarly, the ν parameter should be set to define a trade-off between the information provided by the similar samples and the prior $\mathbf{\Sigma}_0$. A higher ν is needed if few similar patches are available or a large part of the patch is unknown.

Figure 6.2: Ground-truth images used for the experiments presented in Section 6.3.7.

Despite these intuitive insights, the setting of these parameters is not a trivial task and should be the subject of further study. In the perspectives of this work, an in-depth analysis of the role they play in the algorithm and a corresponding thorough setting, are considered.

For this first version and the results presented in Section 6.3.7, the parameters are defined as follows. Because the importance we give to the priors μ_0 and $\mathbf{\Sigma}_0$ is controlled by the relative importance of the value of κ and ν with respect to the diagonal values of the term

$$\sum_{j=1}^{N} \mathbf{\Lambda}^{-1} U_j \mathbf{\Lambda}_j^* U_j, \tag{6.61}$$

(see equations (6.49) and (6.51)), we decide to set κ and ν proportional to the mean value of the diagonal entries of this matrix. The proportionality constant depends on the number of known pixels in the current patch and on the number of similar patches N. If the number of known pixels and that of similar patches are above a threshold, the proportionality constant takes a low value α_L (the same for κ and ν) indicating confidence on the similar patches. Otherwise it takes a high value α_H (also the same for κ and ν). This is of course a simplified and arbitrary procedure, which is in line with the intuitive way of setting these parameters but which can be enhanced.

6.3.5 Summary of the proposed algorithm

The analysis presented in the previous sections leads to a method that is summarized in Algorithm 4, that is an implementation of the Bayesian reconstruction method introduced in Section 6.3. In practice, the algorithm is found to converge after 3 to 4 iterations.

6.3.6 PLE extension to Model (6.22)

The PLE algorithm can be adapted to images following Model (6.22). The *estimation* step is performed using equation (6.25) instead of (6.6) to compute $\tilde{\mathrm{C}}_i^k$, and the class selection equation (6.7) becomes

$$\tilde{k}_i = \arg\min_k \left((\mathrm{z}_i - \boldsymbol{D}_i \tilde{\mathrm{C}}_i^k)^T \mathbf{\Sigma}_{\mathrm{N}_i}^{-1} (\mathrm{z}_i - \boldsymbol{D}_i \tilde{\mathrm{C}}_i^k) + (\tilde{\mathrm{C}}_i^k)^T \mathbf{\Sigma}_k^{-1} \tilde{\mathrm{C}}_i^k + \ln |\mathbf{\Sigma}_k| \right). \tag{6.62}$$

Algorithm 4: Summary of the proposed algorithm.

Decompose the input image Z and $\boldsymbol{D}$ into overlapping patches.
Initialization:
  Compute the covariance matrix of the K classes from synthetic images of edges plus the DCT.
  Project all patches into the K classes using (6.25) and chose the best class using (6.62).
  Compute the first oracle aggregating the estimations of all patches for the chosen class.
Start:
for *it = 1 to maxIts* **do**
  for *all patches not yet restored* **do**
    Find patches similar to the current z_i computing L^2 distance in the oracle.
    Compute MLE of μ_0 and $\boldsymbol{\Sigma}_0$ from similar patches in the oracle.
    Compute $\hat{\mu}$ and $\hat{\boldsymbol{\Sigma}}$ following Algorithm 3.
    Restore the similar patches using (6.25) (colaborative filtering).
  end
  Perform aggregation to restore the image
  The new oracle is equal to the restored image.
end

The rest of the algorithm remains unchanged.

The NLB algorithm may also be extended to the case of images following Model (6.22) taking certain considerations for the *first step*. For the *second step*, the only modification to be done is to replace Equation (6.21) by Equation (6.25) to estimate $\tilde{C}_i$.

6.3.7 Experiments

In this section we summarize the results of a series of experiments conducted to verify the ability of the proposed method to interpolate random missing pixels and denoise the known ones. The interpolation and denoising capacities are evaluated both jointly and separately. The interpolation capacity is evaluated using random masks with different masking rates and the results are compared to those obtained with the PLE method. The denoising capacity is tested adding Gaussian noise with different variance values and the results are compared to those obtained by the NLB algorithm. Finally, the joint interpolation and denoising is tested with images both masked and corrupted by additive Gaussian noise. In this case, the algorithm performance is compared to that of the PLE method only, since NLB was not conceived for pixel interpolation. The ground-truth images for all the tests are shown in Figure 6.2.

Interpolation Random masks with 20%, 50% and 70% of missing pixels are applied to the images in Figure 6.2. The interpolation is computed using the PLE method with the parameters set as indicated in [Yu et al. 2012] (patch size 8×8, $\sigma = 3$, $\varepsilon = 30$). The

	PSNR (dB)					
	barbara		**boat**		**traffic**	
% missing pixels	PLE	HPNLB	PLE	HPNLB	PLE	HPNLB
20	42.98	**45.21**	40.26	**41.62**	34.40	**34.55**
50	43.46	**45.75**	41.10	**41.93**	36.12	**36.16**
70	39.13	**41.38**	37.74	**38.54**	32.97	**33.31**

Table 6.1: Results for the interpolation test with random masks with 20%, 50% and 70% of missing pixels.

	PSNR (dB)					
	extract 1	extract 2	extract 3	extract 4	extract 5	extract 6
HPNLB	**31.58**	**33.72**	**33.61**	**29.69**	**30.55**	**38.70**
PLE	29.01	32.06	32.43	28.55	30.43	37.49

Table 6.2: PSNR values for the extracts shown in Figures 6.3 to 6.5. There are 70% of missing pixels for extracts 1, 2 and 4; 50% for extracts 3 and 5; 20% for extract 6.

parameters for the proposed method are set to: patch size 8×8, $\alpha_H = 1$, $\alpha_L = 0.5$ (α_H and α_L define the values for κ and ν, see Section 6.3.4). The PSNR results are shown in Table 6.1. Figures 6.3 to 6.5 show some extracts of the obtained results and the corresponding difference image with respect to the ground-truth. The PSNR for these extracts are shown in Table 6.2. As it can be verified in the presented extracts, the results obtained by the proposed approach are sharper than the ones obtained by PLE. The improvement is more noticeable comparing the difference images, which keep more structure in the result obtained by PLE. This is quite noticeably in the extracts of barbara's image (Figure 6.3) and in the ropes and cables of the boat image in Figure 6.5. The difference is less noticeable for the lower masking rate as can be seen in Figure 6.5. The gain in PSNR is also important for the larger masking rates.

Denoising The following experiments are conducted in order to compare the denoising ability of the proposed method to that of the NLB algorithm. The experiments are performed with images corrupted with additive Gaussian noise with variance $\sigma^2 = 10, 30, 50$ and 80. For this experiment, there are no unknown pixels to interpolate (the mask $\boldsymbol{D}$ is the identity matrix). From the corresponding algorithms descriptions in Sections 6.2 and 6.3, with a correct parameter setting, the results obtained by these two algorithms should be quite similar. The main difference between them lies in the computation of the mean and covariance matrix of the Gaussian patch model. However, if the prior parameters μ_0 and $\boldsymbol{\Sigma}_0$ are computed in a similar way to what is done in NLB to compute the Gaussian model parameters, and if the parameters κ and ν are large enough to highly prioritize μ_0 and $\boldsymbol{\Sigma}_0$ in equations (6.49) and (6.50), then the Gaussian parameters in both methods will be very close and both methods should lead to very similar results.

That is what we find in practice, with $\alpha_H = \alpha_L = 100$ (α_H and α_L define the values for κ and ν, see Section 6.3.4), as is exemplified in the results summarized in Table 6.3. Figures 6.6 to 6.8 show some extracts that confirm the similarity of the obtained results

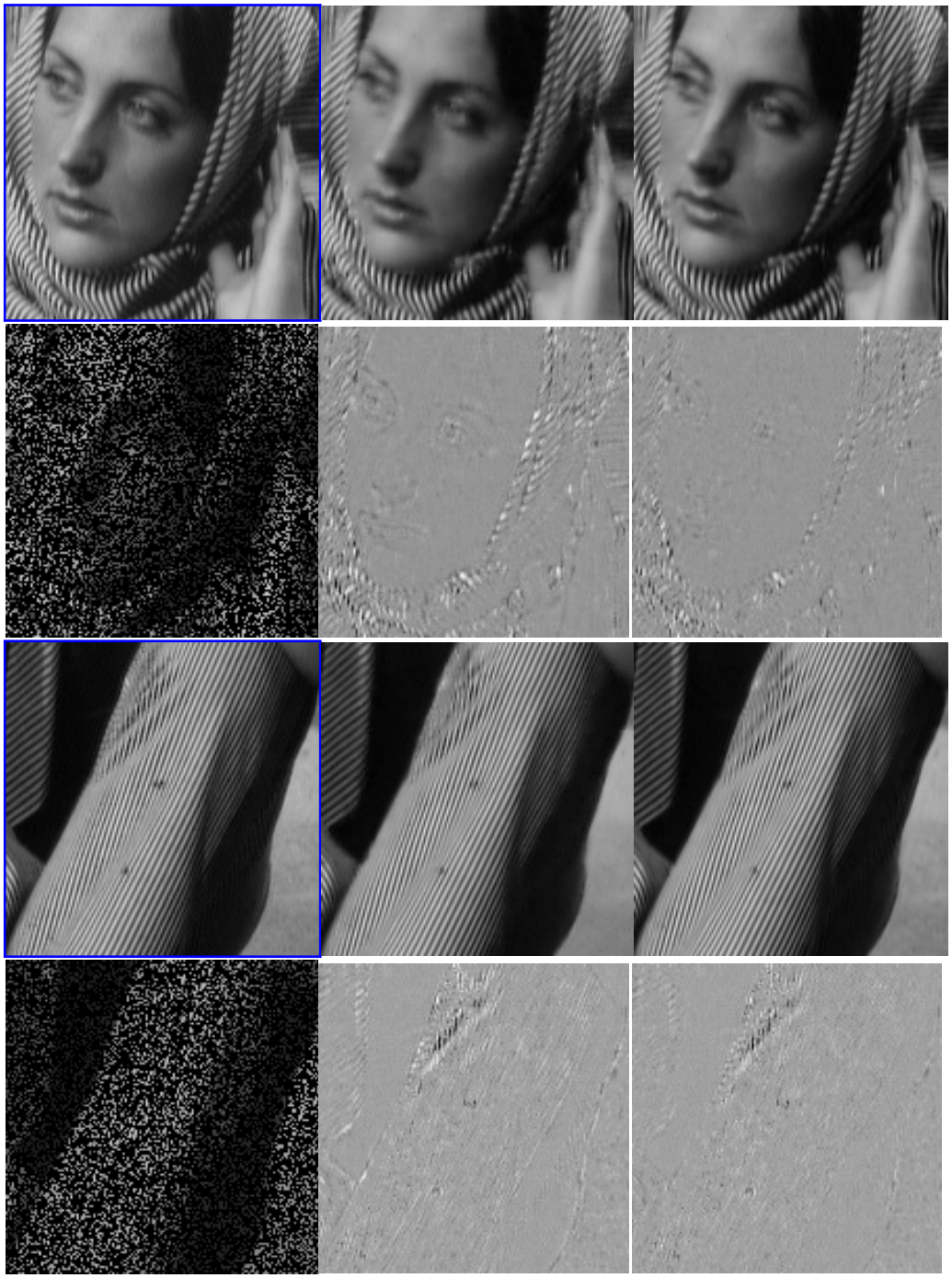


Figure 6.3: **Synthetic data. Interpolation of missing pixels** 70% missing samples. Patch size 8×8. **Left to right:** (first row) Ground-truth, result obtained using PLE, result obtained using the proposed approach, (second row) input image, difference ground-truth - PLE, difference ground-truth - HPNLB. Corresponding PSNR results in Table 6.2. Please see the digital copy for better details reproduction.

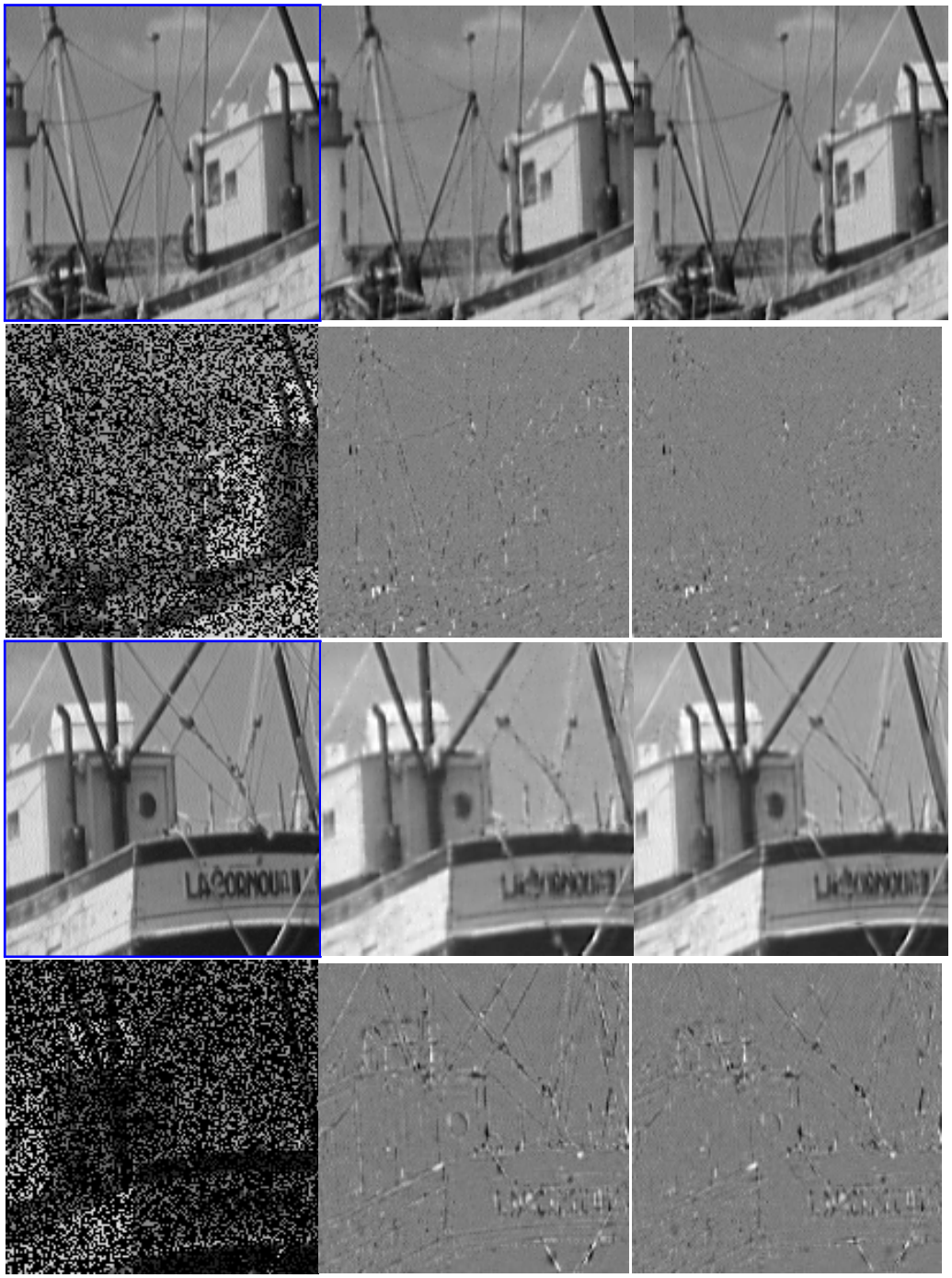


Figure 6.4: **Synthetic data. Interpolation of missing pixels** 50% and 70% missing samples. Patch size 8×8. **Left to right:** (first row) Ground-truth, result obtained using PLE, result obtained using the proposed approach, (second row) input image, difference ground-truth - PLE, difference ground-truth - HPNLB. Corresponding PSNR results in Table 6.2. Please see the digital copy for better details reproduction.

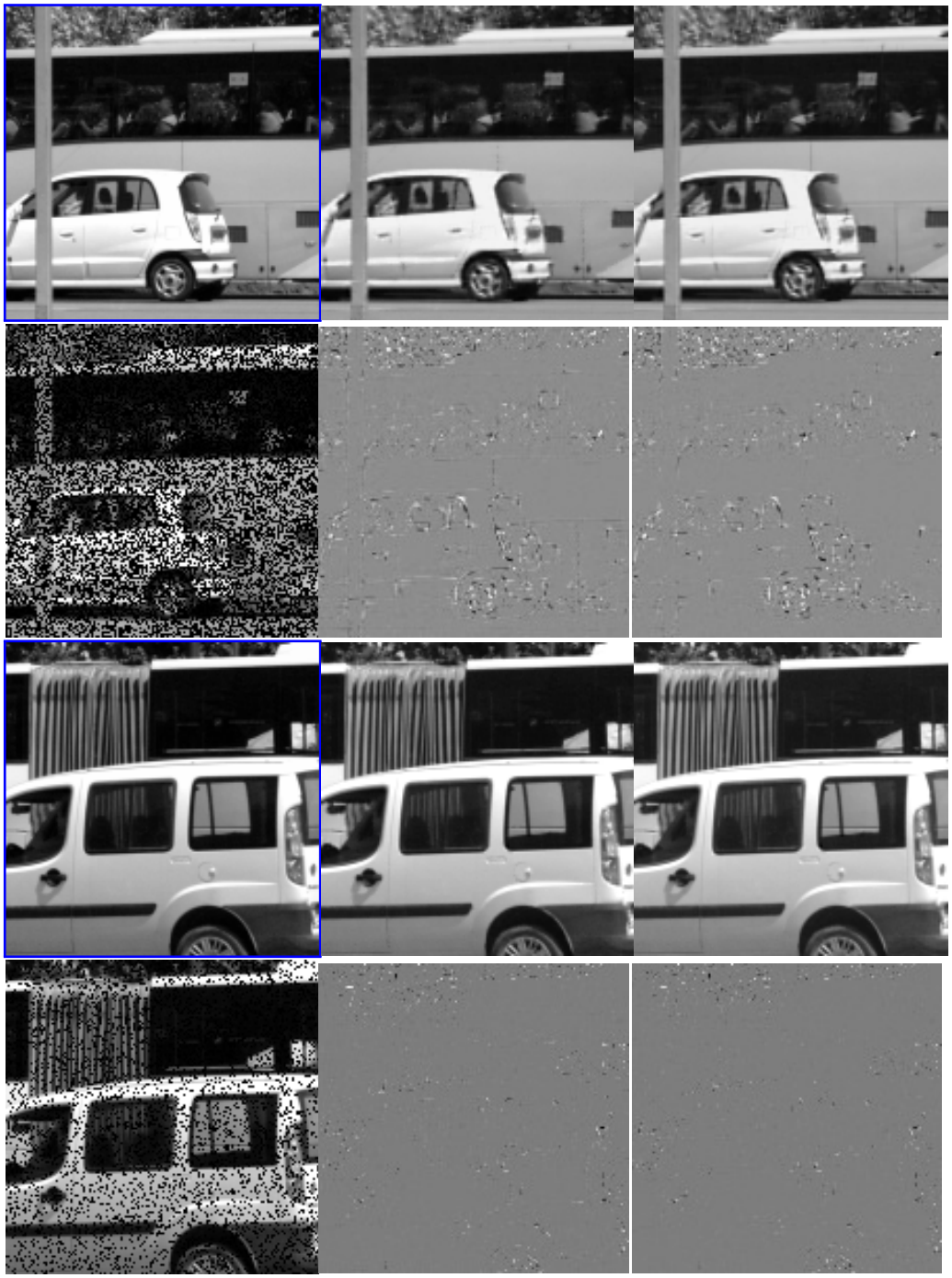


Figure 6.5: **Synthetic data. Interpolation of missing pixels** 50% and 20% missing samples. Patch size 8×8. **Left to right:** (first row) Ground-truth, result obtained using PLE, result obtained using the proposed approach, (second row) input image, difference ground-truth - PLE, difference ground-truth - HPNLB. Corresponding PSNR results in Table 6.2. Please see the digital copy for better details reproduction.

	PSNR (dB)					
	barbara		**boat**		**traffic**	
σ^2	NLB	HPNLB	NLB	HPNLB	NLB	HPNLB
10	40.59	**40.79**	40.03	**40.24**	**40.71**	40.70
30	37.68	**37.98**	36.76	**36.84**	36.97	**37.00**
50	36.34	**36.74**	35.46	**35.56**	35.23	**35.28**
80	35.12	**35.58**	34.32	**34.51**	33.73	**34.06**

Table 6.3: Results for the denoising only test. Images corrupted with additive Gaussian noise with variance 10, 30, 50 and 80. Figures 6.6 to 6.8 show extracts of the obtained results.

	PSNR (dB)					
	extract 1	extract 2	extract 3	extract 4	extract 5	extract 6
HPNLB	**40.27**	**38.88**	**35.18**	**34.05**	37.25	**35.57**
NLB	40.16	38.53	34.93	33.67	**37.29**	35.05

Table 6.4: Denoising only test. Results for the extracts in Figures 6.6 to 6.8. Images corrupted with additive Gaussian noise with variance 10 (extract 1), 30 (extracts 2 and 5), 50 (extract 3) and 80 (extracts 4 and 6). show extracts of the obtained results.

(see Table 6.4 for the PSNR values in these extracts). In order to further emphasize the similarity between $(\mu_0, \mathbf{\Sigma}_0)$ and $(\mu, \mathbf{\Sigma})$ in this case, the proposed method is initialized with the output of the first step of NLB instead of the PLE based initialization described in Section 6.3.5. Notice that this change is possible in this experiment since there are no unknown pixels to interpolate.

We might also think that the Gaussian parameters should be similar for both methods, provided that μ_0 and $\mathbf{\Sigma}_0$ are similar to the Gaussian parameters for NLB, even if κ and ν are small. For instance, since the first term of (6.49) is the addition of the restored patches, we might think that (6.49) should give a result similar to the one obtained by NLB for μ even if κ is small. A similar reasoning can be done for (6.50) and small ν. However, this is not the case in practice, the results obtained by NLB being better than those obtained by the proposed method with small κ and ν. The reason for this is that μ_0 and $\mathbf{\Sigma}_0$, as well as μ and $\mathbf{\Sigma}$ in NLB, are computed from an oracle image resulting from the first restoration step. This restoration includes not only the denoising of each patch, but also an aggregation step that highly improves the final result. Therefore, the contribution of the first term of (6.49) to the computation of $\hat{\mu}$ degrades the result compared to using μ_0 only (i.e. using a large κ).

Combined interpolation and denoising For this experiment, the ground-truth images in Figure 6.2 are corrupted with additive Gaussian noise with variance 10, and a random mask with 20% and 70% of missing pixels. The image restoration is computed using the PLE method with the parameters set as indicated in [Yu et al. 2012] (patch size 8×8, $\sigma = 3$, $\varepsilon = 30$). The parameters for the proposed method are set to: patch size 8×8, $\alpha_H = 1$, $\alpha_L = 0.5$ (α_H and α_L define the values for κ and ν, see Section 6.3.4). Table 6.5

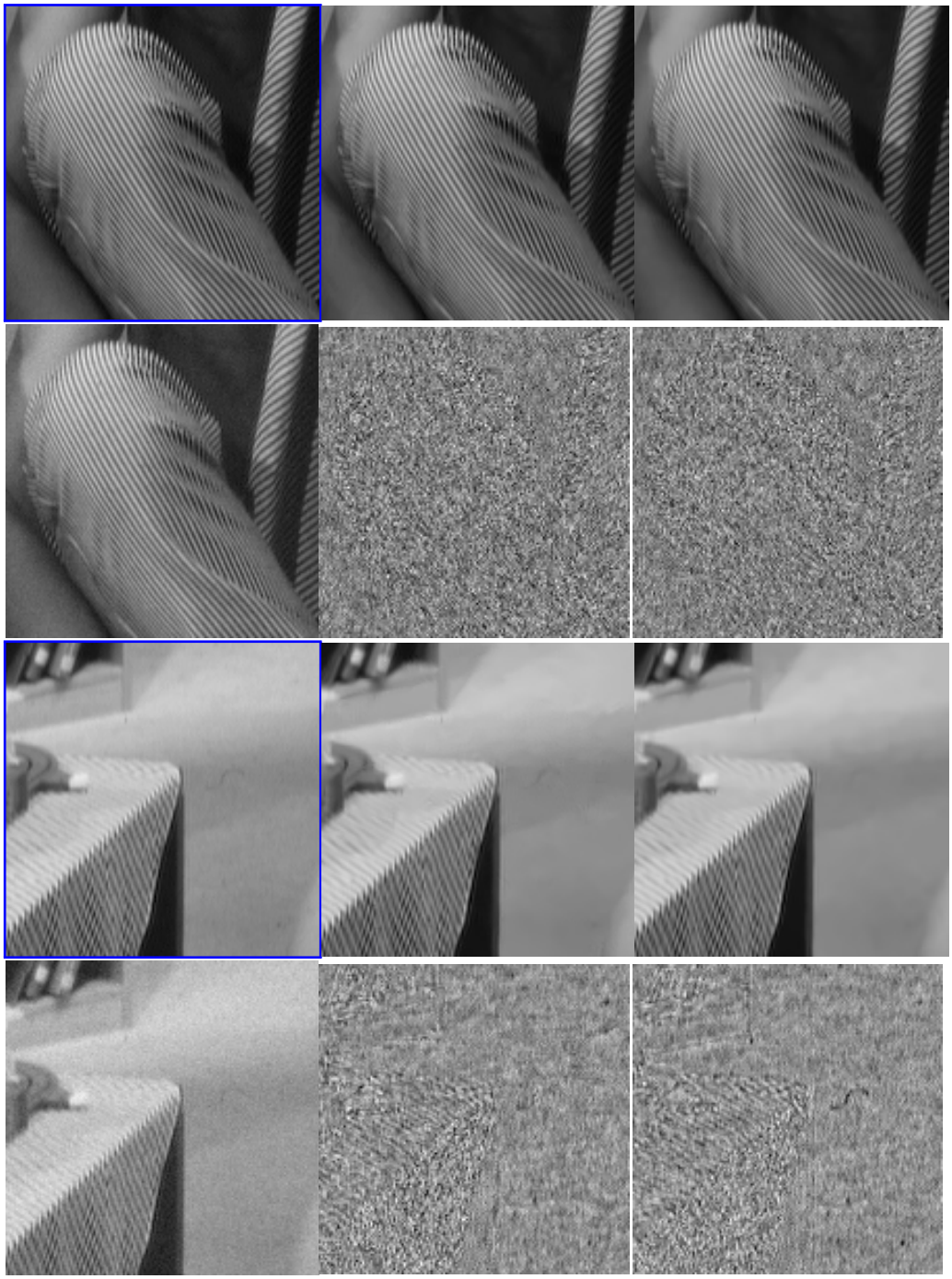


Figure 6.6: **Synthetic data. Denoising** Gaussian additive noise of variance 10 and 30. Patch size 8×8. **Left to right:** (first row) Ground-truth, result obtained using PLE, result obtained using the proposed approach, (second row) input image, difference ground-truth - PLE, difference ground-truth - HPNLB. Corresponding PSNR results in Table 6.4. Please see the digital copy for better details reproduction.

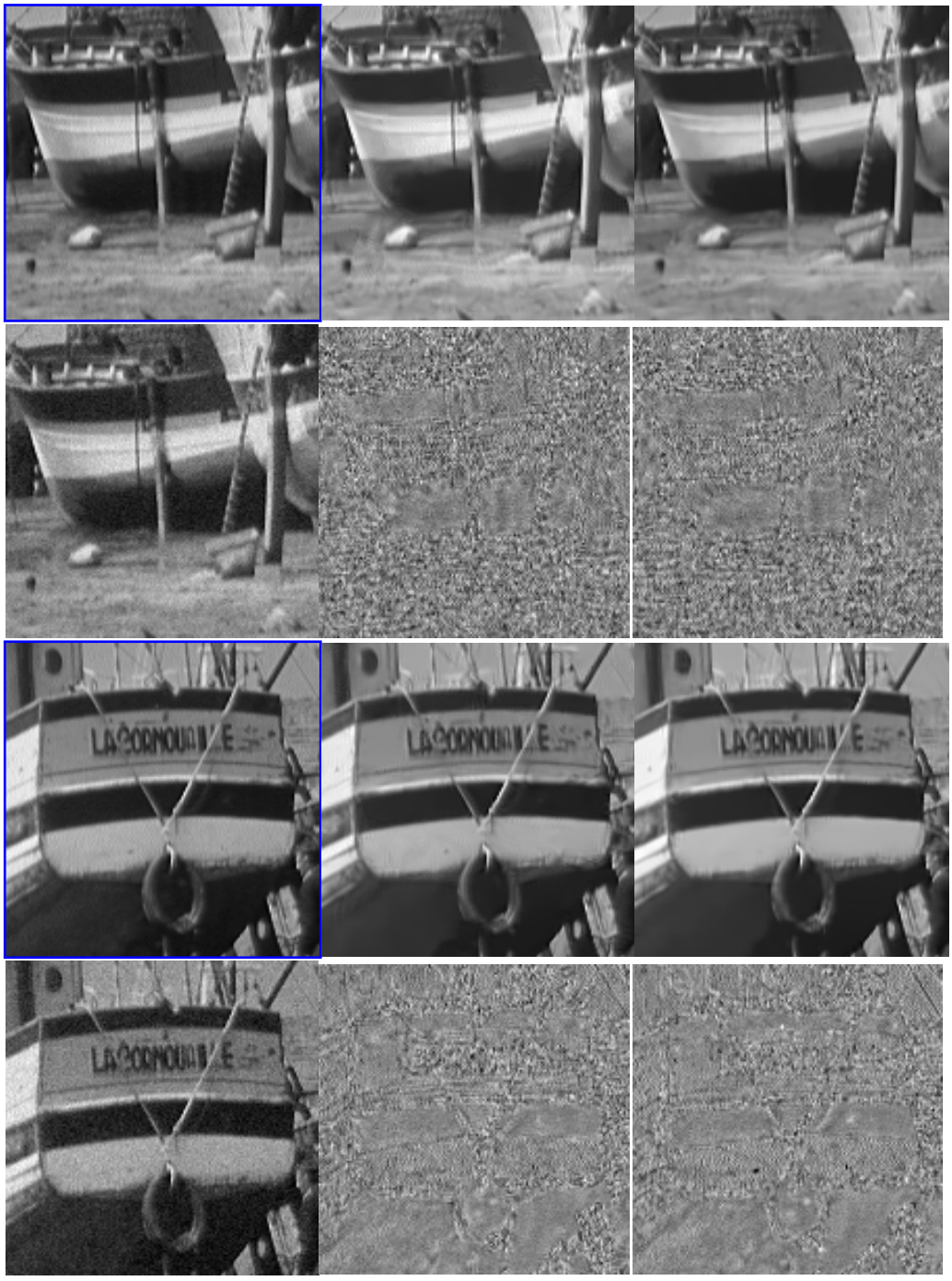


Figure 6.7: **Synthetic data. Denoising** Gaussian additive noise of variance 50 and 80. Patch size 8×8. **Left to right:** (first row) Ground-truth, result obtained using PLE, result obtained using the proposed approach, (second row) input image, difference ground-truth - PLE, difference ground-truth - HPNLB. Corresponding PSNR results in Table 6.4. Please see the digital copy for better details reproduction.

Figure 6.8: **Synthetic data. Denoising** Gaussian additive noise of variance 30 and 80. Patch size 8×8. **Left to right:** (first row) Ground-truth, result obtained using PLE, result obtained using the proposed approach, (second row) input image, difference ground-truth - PLE, difference ground-truth - HPNLB. Corresponding PSNR results in Table 6.4. Please see the digital copy for better details reproduction.

	PSNR (dB)					
	barbara		**boat**		**traffic**	
% missing pixels	PLE	HPNLB	PLE	HPNLB	PLE	HPNLB
20	**37.87**	37.74	37.02	**37.15**	33.30	**33.36**
70	31.46	**32.87**	30.40	**30.99**	25.78	**25.91**

Table 6.5: Results for the denoising and interpolation test. Images corrupted with additive Gaussian noise of variance 10 and random masking with 20% and 70% of missing pixels.

	PSNR (dB)							
	extract 1	extract 2	extract 3	extract 4	extract 5	extract 6	extract 7	extract 8
HPNLB	**37.10**	**36.79**	**33.22**	**30.52**	**30.37**	**27.80**	**28.21**	**26.37**
PLE	37.09	36.57	31.25	28.64	29.87	26.98	26.61	26.07

Table 6.6: PSNR values for the extracts shown in Figures 6.9 to 6.12. Images corrupted with additive noise with variance 10 and random masking with 20% (1-2) and 70% (3-8) missing pixels.

summarizes the PSNR values obtained by each method. Figures 6.9 to 6.12 show the results obtained on some extracts and the corresponding difference images with respect to the ground-truth. Table 6.6 shows the PSNR values for these extracts. For the masking rate of 20%, the results are very similar for both methods. However, a clear difference is observed in the case of 70% masking rate, where the proposed method gives sharper results.

6.4 Application: single image HDR

The idea of using multiple differently exposed images to capture HDR scenes can be traced back to 1856, when the French photographer Gustave Le Gray captured a high dynamic range scene composed by sea and sky, by combining two differently exposed negatives. As explained in Chapter 3, this idea was introduced in digital photography by Mann and Picard in 1995 and several methods followed, proposing different ways to combine the images [Debevec and Malik 1997; Mitsunaga and Nayar 1999; Robertson et al. 2003; Granados et al. 2010].

In the case of a static scene and a static camera, the combination of multiple images is a simple and efficient solution. However, several problems arise when either the camera or the elements in the scene move. As introduced in Chapter 5 for the HDR imaging problem for dynamic scenes and hand-held camera, global alignment techniques must be used to align images and de-ghosting methods must be applied to correct for the artifacts due to object motion. These kind of artifacts are particularly annoying on the fused result.

The multi-image method introduced in Chapter 5 addresses these problems through the use of patches, avoiding the need for image alignment and object motion detection. However, since it uses a reference image, large saturated regions may need to be filled before applying the patch based reconstruction. As will be explained in the following, the single-image method here proposed overcomes this issue.

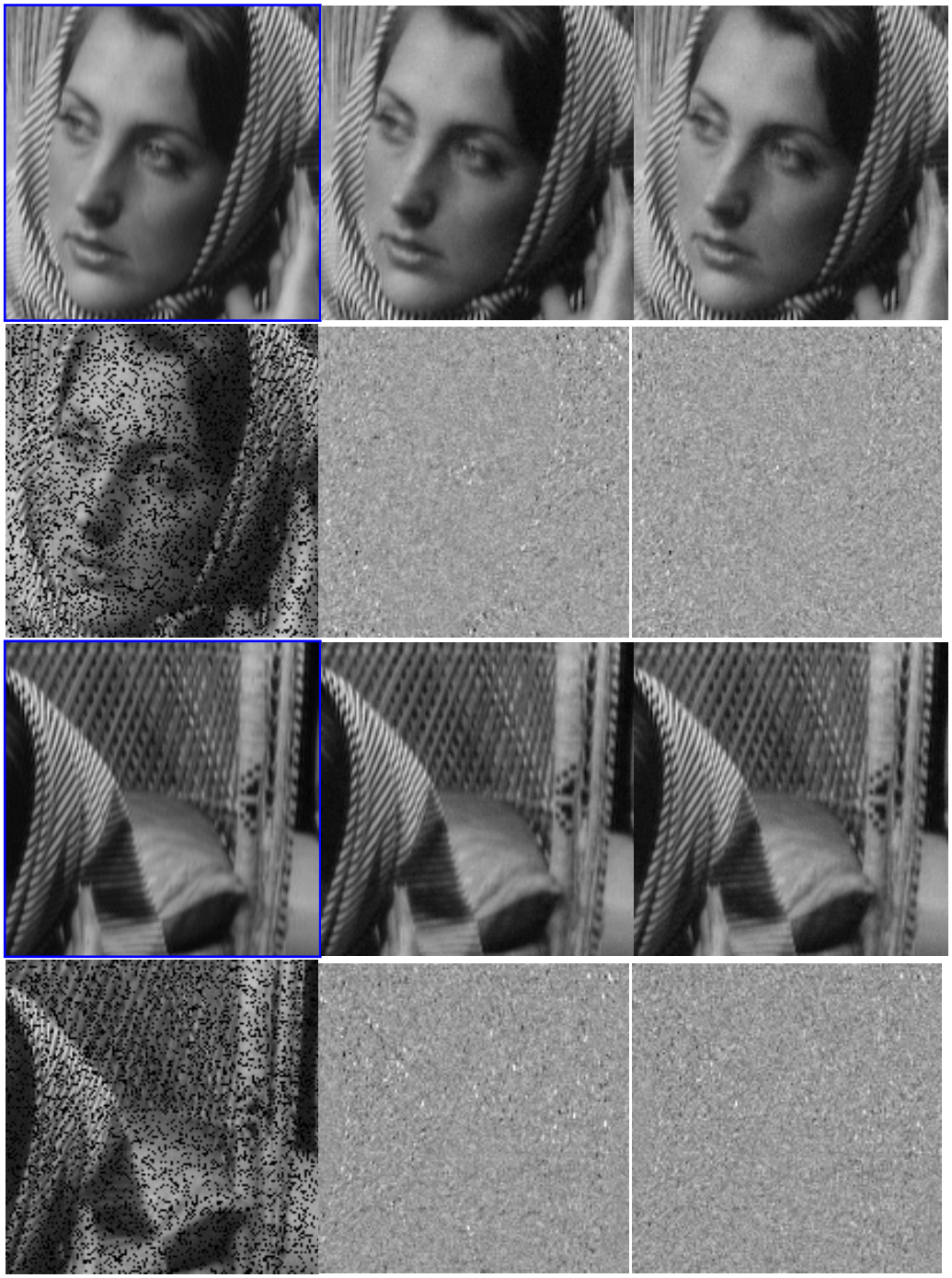


Figure 6.9: **Synthetic data. Combined interpolation and denoising of missing pixels** 20% missing samples + Gaussian noise 10. Patch size 8×8. **Left to right:** (first row) Ground-truth, result obtained using PLE, result obtained using the proposed approach, (second row) input image, difference ground-truth - PLE, difference ground-truth - HPNLB. Corresponding PSNR results in Table 6.6. Please see the digital copy for better details reproduction.

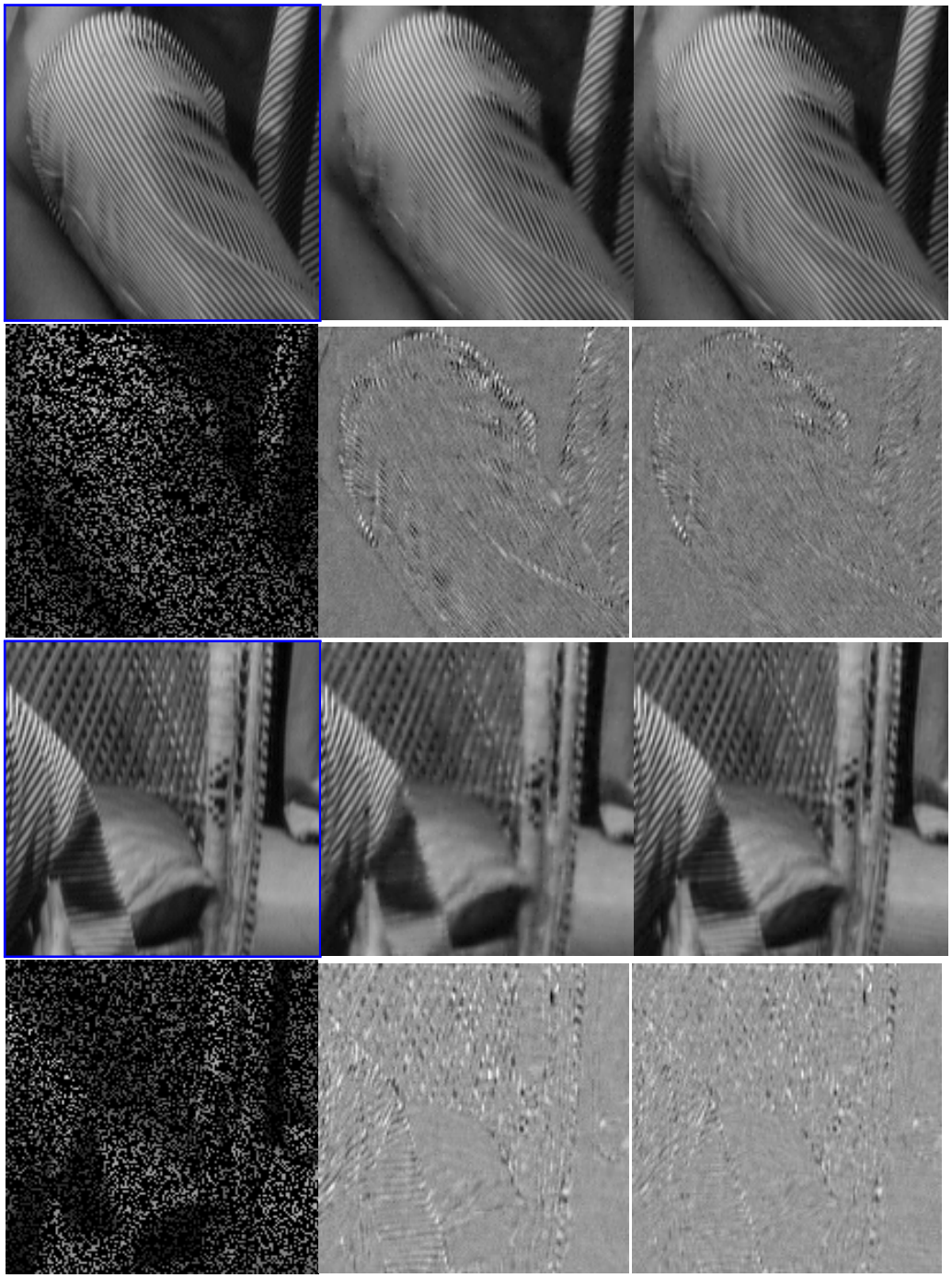


Figure 6.10: **Synthetic data. Combined interpolation and denoising of missing pixels** 70% missing samples + Gaussian noise 10. Patch size 8×8. **Left to right:** (first row) Ground-truth, result obtained using PLE, result obtained using the proposed approach, (second row) input image, difference ground-truth - PLE, difference ground-truth - HPNLB. Corresponding PSNR results in Table 6.6. Please see the digital copy for better details reproduction.

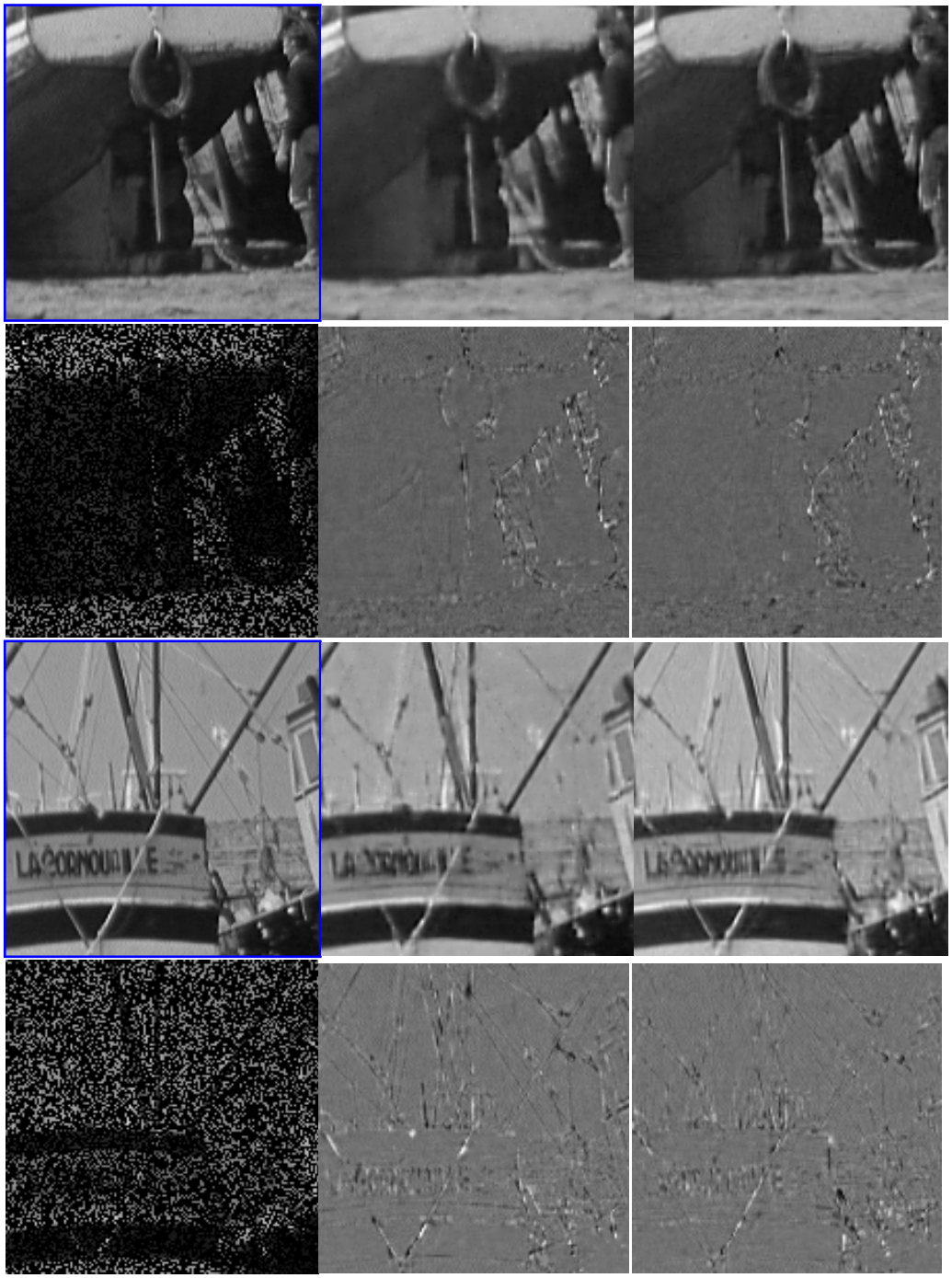


Figure 6.11: **Synthetic data. Combined interpolation and denoising of missing pixels** 70% missing samples + Gaussian noise 10. Patch size 8×8. **Left to right:** (first row) Ground-truth, result obtained using PLE, result obtained using the proposed approach, (second row) input image, difference ground-truth - PLE, difference ground-truth - HPNLB. Corresponding PSNR results in Table 6.6. Please see the digital copy for better details reproduction.

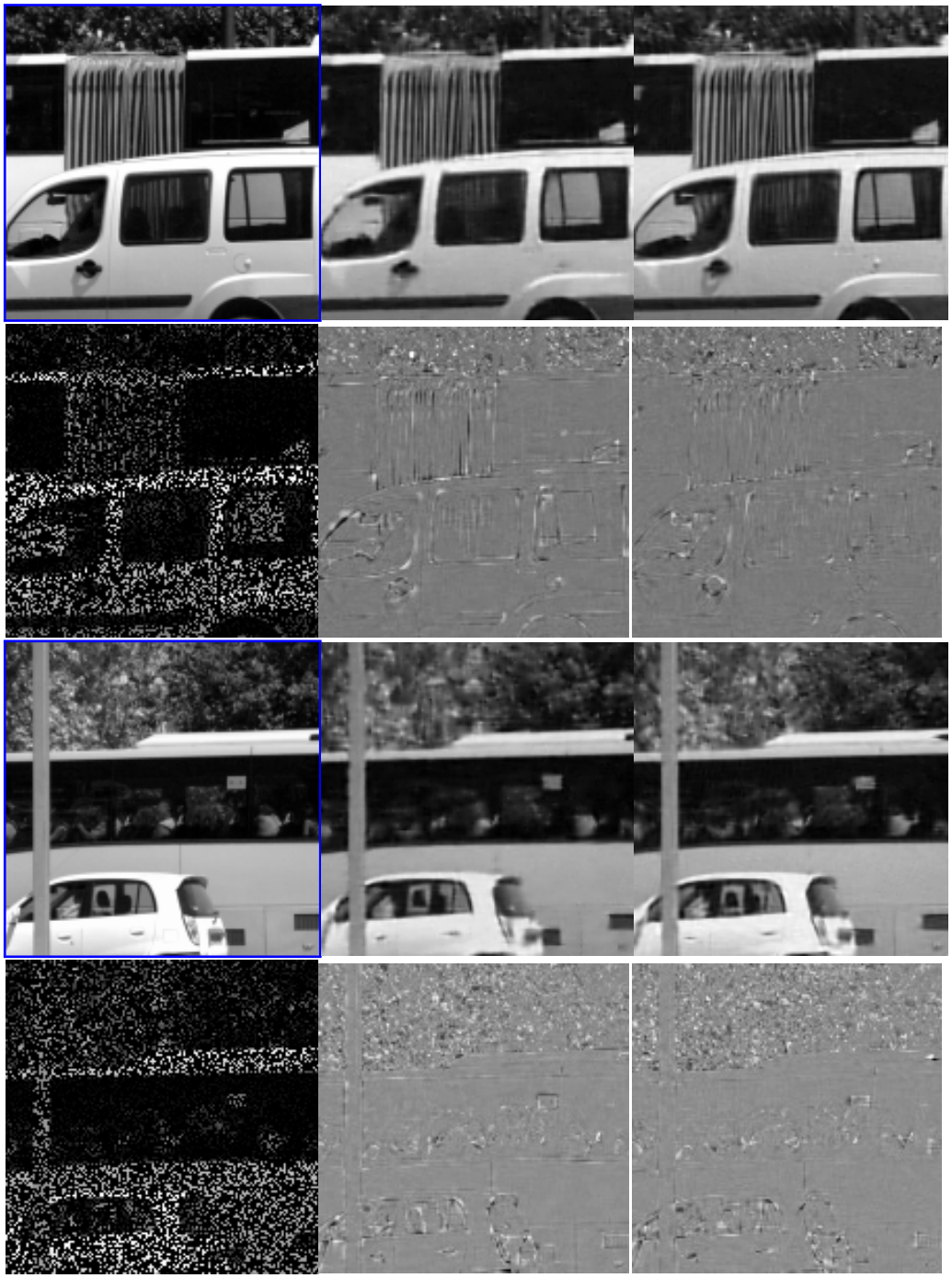


Figure 6.12: **Synthetic data. Combined interpolation and denoising of missing pixels** 70% missing samples + Gaussian noise 10. Patch size 8×8. **Left to right:** (first row) Ground-truth, result obtained using PLE, result obtained using the proposed approach, (second row) input image, difference ground-truth - PLE, difference ground-truth - HPNLB. Corresponding PSNR results in Table 6.6. Please see the digital copy for better details reproduction.

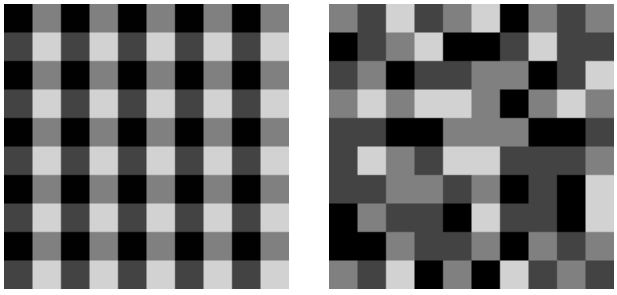


Figure 6.13: Regular (left) and non-regular (right) optical masks for an example of 4 different filters.

The concept of spatially varying pixel exposures (SVE) proposed by Nayar and Mitsunaga [2000] enables to capture a very large range of exposures in only one shot while avoiding multi-imaging limitations. In this work, we propose a novel approach to generate HDR images from a single shot acquired with spatially varying pixel exposures. Drawing on a precise modeling of the camera acquisition noise, we make use of the new image restoration technique introduced in Section 6.3. The proposed method reconstructs the irradiance image by simultaneously estimating missing pixels and denoising existing ones, showing considerable improvements over existing approaches.

This section is organized as follows. Section 6.4.1 presents a summary of the previous work on single shot HDR imaging. In Section 6.4.2 we present the SVE acquisition model. Section 6.4.3 is devoted to the irradiance reconstruction problem and the proposed solution. A summary of the performed experiments is presented in Section 6.4.4. Conclusions are presented in Section 6.5.

6.4.1 Previous work

An alternative to HDR from multiple frames was introduced by Nayar and Mitsunaga [2000]. They propose to perform HDR imaging form a single image using spatially varying pixel exposures. An optical mask with spatially varying transmittance (see Figure 6.13) is placed adjacent to a conventional image sensor, thus controlling the amount of light that reaches each pixel. This gives different exposure levels to the pixels according to the given transmittance pattern, allowing a single shot to capture an increased dynamic range compared to that of the conventional sensor.

The greatest advantage of this acquisition method is that it allows HDR imaging from a single image, thus avoiding the need for alignment and motion estimation, which is the main drawback of the classical multi-image approach. Another advantage is that the saturated pixels are not organized in large regions. Indeed, some recent multi-image methods tackle the camera and objects motion problems by taking a reference image and then estimating motion relative to this frame or by recovering information from other frames through local comparison with the reference [Sen et al. 2012; Aguerrebere et al. 2013] (Chapter 5). A problem encountered by this approach is the need for inpainting saturated and underexposed regions in the reference frame, since the information is completely lost in those areas. The SVE acquisition strategy prevents from having large saturated regions to inpaint. In general, all scene regions are sampled by at least one of the exposures thus

simplifying the inpainting problem.

The main drawback of the SVE acquisition is that, unlike the multi-image approach where all pixels are assumed to be correctly exposed in at least one of the input images, for the brighter and darker regions of the scene some exposure levels will be either too high or too low and the corresponding pixels will be under or over exposed. Hence, those pixels are unknown and need to be somehow reconstructed. Figure 6.14 illustrates this problem. It shows an example of an HDR scene and the mask of known and unknown pixel values of a single shot of the scene using SVE. Known pixels (white) are the correctly exposed pixels and unknown (black) pixels are those either under or over exposed. Moreover, noise reduction is of particular importance in this kind of acquisition setup since the pixels of the lowest exposures tend to be quite noisy (mostly in dark regions) thus producing images with high noise levels.

Hirakawa and Simon [2011] argue that different sensitivities are already implied by the different translucencies of the three color filters in a regular Bayer Pattern. They propose a clever demosaicking-inspired algorithm to jointly perform demosaicking and HDR imaging from a single shot, with specially tailored color-filter translucencies. Moreover, they propose to use an external optical filter in order to increase the differences between the translucencies present in the color filters. Despite this solution may be, in certain conditions, more limited than the SVE strategy, it is remarkably practical (no need to modify the image sensor) and therefore very interesting from a feasibility point of view.

In the approach proposed by Nayar and Mitsunaga [2000], the varying exposures follow a regular pattern as shown in Figure 6.13. Two methods are proposed to reconstruct the under and over exposed pixels. The so called aggregation approach consists in averaging the local irradiance values produced by the correctly exposed pixels. The interpolation approach consists in using a bi-cubic interpolation to simultaneously retrieve the unknown pixels and denoise the known ones. A generalization of this kind of pixel varying acquisition, and its application to high dynamic range and multi-spectral imaging is presented in [Yasuma et al. 2010].

Motivated by the aliasing problems of regular sampling patterns, Schöberl et al. [2012a] propose to use spatially varying exposures in a non-regular pattern. Figure 6.13 shows examples of both acquisition patterns: regular and random. The reconstruction of the irradiance image is then performed using the frequency selective extrapolation algorithm [Seiler and Kaup 2010] which iteratively generates a sparse model for each image patch as a weighted superposition of the two-dimensional Fourier basis functions. Moreover, Schöberl et al. [2012b] present a practical methodology for the construction of a spatially varying exposures mask with a non-regular pattern.

In [Aguerrebere et al. 2014a], we proposed a new HDR imaging technique that combines the SVE acquisition strategy with an irradiance reconstruction using an extension of the PLE method adapted to the noise present in raw digital images presented in Section 6.3.6. That is, noise with variable variance dependent on the irradiance on each pixel (c.f. Chapter 2). The raw images are first normalized to the irradiance domain (division by the corresponding SVE gains) and the modified PLE method is applied to the set of image patches. As will be shown in Section 6.4.4, the method presented in this section outperforms the one introduced in [Aguerrebere et al. 2014a].

Figure 6.14: Example of the acquisition of an HDR scene using spatially varying pixel exposures. **Left:** Tone mapped HDR scene restored from the raw image. **Right top:** Raw image with spatially varying exposure levels. **Right bottom:** Mask of correctly exposed pixels (white) and under or over exposed pixels (black).

In this work, we propose to use the HPNLB method to reconstruct the irradiance information of a scene from a single shot acquired with spatially varying pixel exposures following a random pattern. We take advantage of the GMM, which has been proven accurate at representing natural image patches, to reconstruct the unknown pixels and denoise the correctly exposed ones. The new approach improves the irradiance reconstruction with respect to the previous approaches.

6.4.2 Spatially varying exposure acquisition model

As presented in [Nayar and Mitsunaga 2000; Yasuma et al. 2010; Schöberl et al. 2012b], an optical mask with spatially varying transmittance can be placed adjacent to a conventional image sensor to give different exposure levels to the pixels. This optical mask does not change the acquisition process of the sensor, whether using a conventional CCD or CMOS sensor. As introduced in Chapter 2, the main noise sources for this kind of sensors are: the Poisson photon shot noise, which can be approximated by a Gaussian distribution with equal mean and variance; the thermally generated readout noise, which is modeled as an additive Gaussian distributed noise; the spatially varying gain given by the photo response non uniformity (PRNU); dark currents and quantization noise. Thus the following noise model for the non saturated raw pixel value Z_p at position p is applicable for this new acquisition method[2]:

$$\mathrm{Z}_p \sim \mathcal{N}(g o_p a_p \tau C_p + \mu_R, g^2 o_p a_p \tau C_p + \sigma_R^2), \tag{6.63}$$

where g is the camera gain, o_p is the variable gain due to the optical mask, a_p models the PRNU factor, τ is the exposure time, C_p is the irradiance reaching pixel p, μ_R and σ_R^2 are

[2]Some noise sources not modeled in Chapter 2, such as blooming, might have a considerable impact in the SVE acquisition strategy and should be considered in a more accurate image modeling.

the readout noise mean and variance. Dark currents and quantization noise are neglected (c.f. Chapter 2).

Two main aspects must be defined for the SVE acquisition strategy. One is the number of different filters to be used, i.e. the different exposure levels to capture. This is related to the problem of how many exposure times should be used in the classical HDR acquisition strategy. The solution to this problem depends on the scene. Since the acquisition using SVE uses an a priori fixed optical mask, the number of different exposures is fixed. In general, 2 to 4 images are used for HDR imaging. An optical mask with 4 different exposure levels appears a reasonable choice [Nayar and Mitsunaga 2000].

The second choice is whether the spatial distribution of the different filters is done randomly or with a regular pattern. This determines the way the scene irradiance is sampled: regular or random sampling. Figure 6.13 shows examples of the two sampling strategies. This point is important in the acquisition strategy since, due to unknown under and over exposed pixels, some regions of the image will almost certainly be sub-sampled and some kind of interpolation will be needed to retrieve these pixels values. If the sampling pattern is regular, aliasing artifacts will appear due to the characteristics of the spectrum of the pattern (delta functions at the sampling frequencies). On the contrary, the spectrum of a random pattern is concentrated in a single delta and has negligible values for the rest of the frequencies, thus avoiding aliasing. This fact led us to choose a random pattern to perform the acquisition.

6.4.3 Irradiance reconstruction

In order to reconstruct the dynamic range of the scene we need to solve an inverse problem, that is, to find the irradiance values from the input pixel values. Several widely known methods solve image inverse problems decomposing the image into patches so as to take advantage of accurate models developed to represent patches. In Section 6.3 we introduced a new Bayesian reconstruction algorithm which is able to interpolate missing pixels and denoise the existing ones making use of a Gaussian prior for image patches. This method is well adapted to the problem of the reconstruction of the dynamic range of an image acquired using the SVE acquisition strategy. In the following we present the details of how to apply this method to the reconstruction of the irradiance map from such an image.

The inverse problem

The problem we want to solve is that of estimating the irradiance image C from the input raw image Z, knowing the exposure levels and the camera parameters. Let us consider Y_p the normalization of the input pixel Z_p to the irradiance domain

$$\mathrm{Y}_p = \frac{\mathrm{Z}_p - \mu_R}{g o_p a_p \tau}. \tag{6.64}$$

We take into account the effect of saturation and under-exposure by introducing the exposure degradation factor D_p given by

$$\mathrm{D}_p = \begin{cases} 1 & \text{if } \mu_R < \mathrm{Z}_p < z_{sat}, \\ 0 & \text{otherwise} \end{cases} \tag{6.65}$$

with z_{sat} equal to the pixel saturation value, thus eliminating the under or overexposed pixels. From (6.63), Y_p can be modeled as

$$\mathrm{Y}_p \sim \mathcal{N}\left(\mathrm{D}_p C_p, \frac{g^2 o_p a_p \tau \mathrm{D}_p C_p + \sigma_R^2}{(g o_p a_p \tau)^2}\right). \tag{6.66}$$

Notice that (6.66) is the distribution of Y_p for a given D_p, since D_p is itself a random variable that depends on Z_p. The exposure degradation factor must be included in (6.66) since the variance of the over or under exposed pixels no longer depends on the irradiance C_p but is only due to the readout noise σ_R^2.

Then the problem of irradiance estimation can be stated as retrieving C from the image Y, which implies denoising the known Y_p pixel values ($\mathrm{D}_p = 1$) and estimating the completely unknown ones ($\mathrm{D}_p = 0$).

Hyperprior Non Local Bayes for irradiance reconstruction

In order to reconstruct C from the irradiance image Y we make use of the Bayesian restoration method introduced in Section 6.3, which is adapted to the noise present in the raw irradiance values given by (6.66).

Patch model The irradiance image Y is decomposed into overlapping patches Y_i of size $\sqrt{N} \times \sqrt{N}$, $i = 1, \ldots, I$ with I the number of patches in the image. From (6.66), each patch Y_i taken as a column vector of size $N \times 1$ can be modeled according to

$$\mathrm{Y}_i = \boldsymbol{D}_i \mathrm{C}_i + \boldsymbol{\Sigma}_{\mathrm{N}_i}^{1/2} \varepsilon_i. \tag{6.67}$$

Here, the degradation operator $\boldsymbol{D}_i$ is a $N \times N$ diagonal matrix with the diagonal entries equal to the degradation image $\boldsymbol{D}$ restricted to the patch i, C_i is the patch on the irradiance image we seek to estimate, $\boldsymbol{\Sigma}_{\mathrm{N}_i}$ is a $N \times N$ diagonal matrix with the j-th diagonal entry given by

$$(\boldsymbol{\Sigma}_{\mathrm{N}_i})_j = \frac{g^2 o_p a_p \tau (\boldsymbol{D}_i \mathrm{C}_i)_j + \sigma_R^2}{(g o_p a_p \tau)^2}, \tag{6.68}$$

where $(\boldsymbol{D}_i \mathrm{C}_i)_j$ is the j-th element of vector $\boldsymbol{D}_i \mathrm{C}_i$ and ε_i is a Gaussian noise with zero mean and identity covariance matrix. Notice that the degradation operator $\boldsymbol{D}_i$ is actually a random variable, but we consider it here as a determinist known mask in order to be able to apply the reconstruction method introduced in Section 6.3. A Gaussian prior is chosen to describe image patches, parametrized by its mean μ_i and covariance matrix $\boldsymbol{\Sigma}_i$.

Patch reconstruction The patch reconstruction is performed using the Bayesian method introduced in Section 6.3. Hence, the restored patch $\tilde{\mathrm{C}}_i$ is computed as

$$\tilde{\mathrm{C}}_i = \boldsymbol{\Sigma}_i \boldsymbol{D}_i^T (\boldsymbol{D}_i \boldsymbol{\Sigma}_i \boldsymbol{D}_i^T + \boldsymbol{\Sigma}_{\mathrm{N}_i})^{-1} (\mathrm{y}_i - \boldsymbol{D}_i \mu_i) + \mu_i, \tag{6.69}$$

with the observed noisy irradiance patch y_i given by

$$\mathrm{y}_i = \frac{\mathrm{z}_i - \mu_R}{g o_p a_p \tau}. \tag{6.70}$$

The model parameters μ_i and $\mathbf{\Sigma}_i$ are computed as detailed in Section 6.3.3, and an iterative procedure is performed to improve the accuracy of the parameters estimation as summarized in Section 6.3.5. The proposed approach is summarized in Algorithm 5.

Algorithm 5: Summary of the proposed single-image HDR imaging method

1 Compute the irradiance image Y from the input image Z using (6.64).
2 Compute the degradation mask $\boldsymbol{D}$ from the input image Z using (6.65).
3 Decompose Y and $\boldsymbol{D}$ into overlapping patches.
4 Use Algorithm 4 to compute the irradiance map.

PLE for irradiance reconstruction

The extension of the PLE method to the patch model 6.67 presented in Section 6.3.6 can also be applied to the irradiance reconstruction from a single image captured using the SVE strategy. This work was published in [Aguerrebere et al. 2014a] and is available at the project site[3].

The considered patch model is given by 6.67, and the patch space is represented by K Gaussian distributions or classes, parametrized by their means μ_k and covariance matrices $\mathbf{\Sigma}_k$, $k = 1, \ldots, K$. Each patch C_i is assumed to be drawn independently from one of these K Gaussians, whose probability density functions are given by

$$p(\mathrm{C}) = \frac{1}{(2\pi)^{N/2}|\mathbf{\Sigma}_k|^{1/2}} \exp\left(-\frac{1}{2}(\mathrm{C}-\mu_k)^T\mathbf{\Sigma}_k^{-1}(\mathrm{C}-\mu_k)\right). \tag{6.71}$$

The method consists in two steps which are alternately iterated: the patch reconstruction and the class update steps.

Patch reconstruction In the patch reconstruction step the classes parameters $(\mu_k, \mathbf{\Sigma}_k)$ are assumed to be known. The patch is reconstructed according to

$$\tilde{\mathrm{C}}_i^k = \mathbf{\Sigma}_k\boldsymbol{D}_i^T(\boldsymbol{D}_i\mathbf{\Sigma}_k\boldsymbol{D}_i^T + \mathbf{\Sigma}_{\mathrm{N}_i})^{-1}(\mathrm{y}_i - \boldsymbol{D}_i\mu_k) + \mu_k, \tag{6.72}$$

for all classes $k = 1, \ldots, K$. Notice that 6.72 is equal to 6.69 with $(\mu_k, \mathbf{\Sigma}_k)$ instead of $(\mu_i, \mathbf{\Sigma}_i)$. Then the best representing class $\tilde{k}_i$ for each patch is chosen as

$$\tilde{k}_i = \arg\min_k \left((\mathrm{y}_i - \boldsymbol{D}_i\tilde{\mathrm{C}}_i^k)^T\mathbf{\Sigma}_{\mathrm{N}_i}^{-1}(\mathrm{y}_i - \boldsymbol{D}_i\tilde{\mathrm{C}}_i^k) + (\tilde{\mathrm{C}}_i^k)^T\mathbf{\Sigma}_k^{-1}\tilde{\mathrm{C}}_i^k + \ln|\mathbf{\Sigma}_k|\right). \tag{6.73}$$

Class update Once all patches have been assigned a class through 6.73, the class parameters $(\mu_k, \mathbf{\Sigma}_k)$ are updated computing their MLE estimators

$$\tilde{\mu}_k = \frac{1}{|S_k|}\sum_{j\in S_k}\tilde{\mathrm{C}}_j, \quad \tilde{\mathbf{\Sigma}}_k = \frac{1}{|S_k|}\sum_{j\in S_k}(\tilde{\mathrm{C}}_j - \tilde{\mu}_k)(\tilde{\mathrm{C}}_j - \tilde{\mu}_k)^T, \tag{6.74}$$

[3] `http://perso.telecom-paristech.fr/~gousseau/single_shot_hdr/`

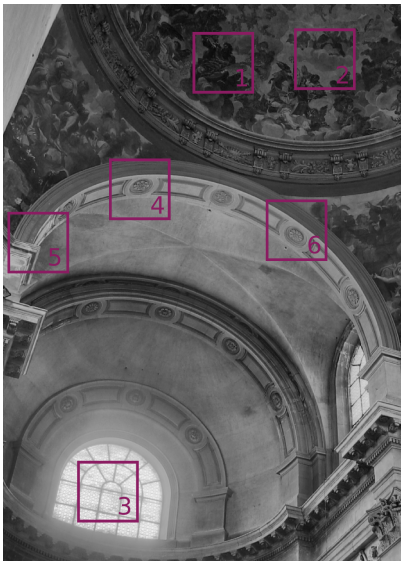


Figure 6.15: Tone mapped version of the ground-truth image used for the experiments in Section 6.4.4. Image from [Hasinoff et al.]. The results for the extracts are shown in Figures 6.16 to 6.21.

with S_k the set of all patches assigned to class k and $|S_k|$ its cardinality. The covariance matrix $\tilde{\mathbf{\Sigma}}_k$ may not be well conditioned as a result, for example, of a small number of patches in the class. For this reason a regularization term ε is added to $\tilde{\mathbf{\Sigma}}_k$ to ensure the correct inversion of the matrix [Yu et al. 2012].

The method is initialized as explained in Section 6.1.1. As mentioned in Section 6.1, the input image is decomposed into regions of size 128×128 and the proposed approach is applied to each region separately. Regions are half-overlapping to avoid boundary effects. Because the image content is more coherent semi-locally than globally, this treatment allows for a better reconstruction with a fixed number of classes K. This semi-local treatment is especially important in the case of HDR images, where the considered dynamic range may be very high and the number of classes needed to represent the image treated as a whole would be very large. As in Yu et al. [2012], we use $K = 20$ in all our experiments. The algorithm is found to converge in 3 to 4 iterations.

6.4.4 Experiments

In this section we present the results obtained in a series of experiments conducted to verify the capacity of the proposed method to retrieve the complete irradiance map from a single image acquired using spatially varying pixel exposures. The experiments are performed using both synthetic and real data. The performance is compared to that obtained using the previous methods by Schöberl et al. [2012a], Nayar and Mitsunaga [2000], and the PLE method adapted to the case of noise with variable variance [Aguerrebere et al. 2014a], briefly introduced in the previous section, and hereafter called PLEV.

	PSNR (dB)					
	1 (Fig. 6.16)	2 (Fig. 6.17)	3 (Fig. 6.18)	4 (Fig. 6.19)	5 (Fig. 6.20)	6 (Fig. 6.21)
HPNLB	**33.08**	**33.87**	22.95	**35.10**	**36.80**	**35.66**
PLEV	29.65	30.82	22.77	33.99	36.42	34.73
Schöberl et al.	30.38	31.16	21.39	30.04	32.84	31.02
Nayar and Mitsunaga	29.39	30.10	**23.24**	25.83	30.26	26.90

Table 6.7: PSNR values for the extracts shown in Figures 6.16 to 6.21.

Synthetic data

Experiments using synthetic data are carried out in order to compare the reconstruction obtained by the proposed and previous methods against a ground-truth, which is not possible (or highly prone to errors) using real data. For this purpose, sample images are generated according to Model (6.63) using the HDR image in Figure 6.15 as ground-truth. Both a random and a regular pattern with four equiprobable exposure levels $o = \{1, 8, 64, 512\}$ are simulated. The exposure time is set to $\tau = 1/200$ seconds and the camera parameters are those of a Canon 7D camera set to ISO 200 ($g = 0.87$, $\sigma_R^2 = 30$, $\mu_R = 2048$, $\mathrm{v}_{\mathrm{sat}} = 15000$). The parameters for the proposed method are set to: patch size 8×8, $\alpha_H = 10$, $\alpha_L = 5$ (α_H and α_L define the values for κ and ν, see Section 6.3.4). For the PLEV method, the patch size is set to 8×8 and $\varepsilon = 0.1$. The parameters for the method by Schöberl et al. were set to obtain the best PSNR value.

Figures 6.16 to 6.21 show extracts of the results obtained by the proposed method, by PLEV and by Schöberl et al. for the random pattern and the results obtained by the bicubic interpolation proposed by Nayar and Mitsunaga using the regular pattern. Together with the results we show the corresponding input images and the difference of each result with respect to the ground-truth image. Table 6.7 shows the PSNR values obtained in each extract by each method.

The extract shown in Figure 6.16 presents a quite difficult reconstruction challenge since it corresponds to a very dark region of the scene (the high ceiling that receives almost no light, see Figure 6.15) and thus almost all pixels are either under-exposed, and thus unknown, or close to become under-exposed, and thus with a very low signal-to-noise ratio. For that extract, 51% of the pixels are unknown and the other known half is very noise. Nevertheless, the proposed approach manages to produce a quite correct reconstruction, both visually and in terms of PSNR. Notice that even if the PLEV method manages to reconstruct the missing pixels, the reconstruction is quite smooth and the denoising results are not satisfactory. A similar result can be seen in Figure 6.17.

The difference in the results is less obvious in the rest of the extracts, which correspond to better lit areas. Nevertheless, the results obtained by the proposed approach remain sharper than those produced by PLEV, as is also reflected in the PSNR values.

Real data

The feasibility of the SVE random pattern has been shown in [Schöberl et al. 2012b] and that of the SVE regular pattern in [Yasuma et al. 2010]. Nevertheless, these acquisition

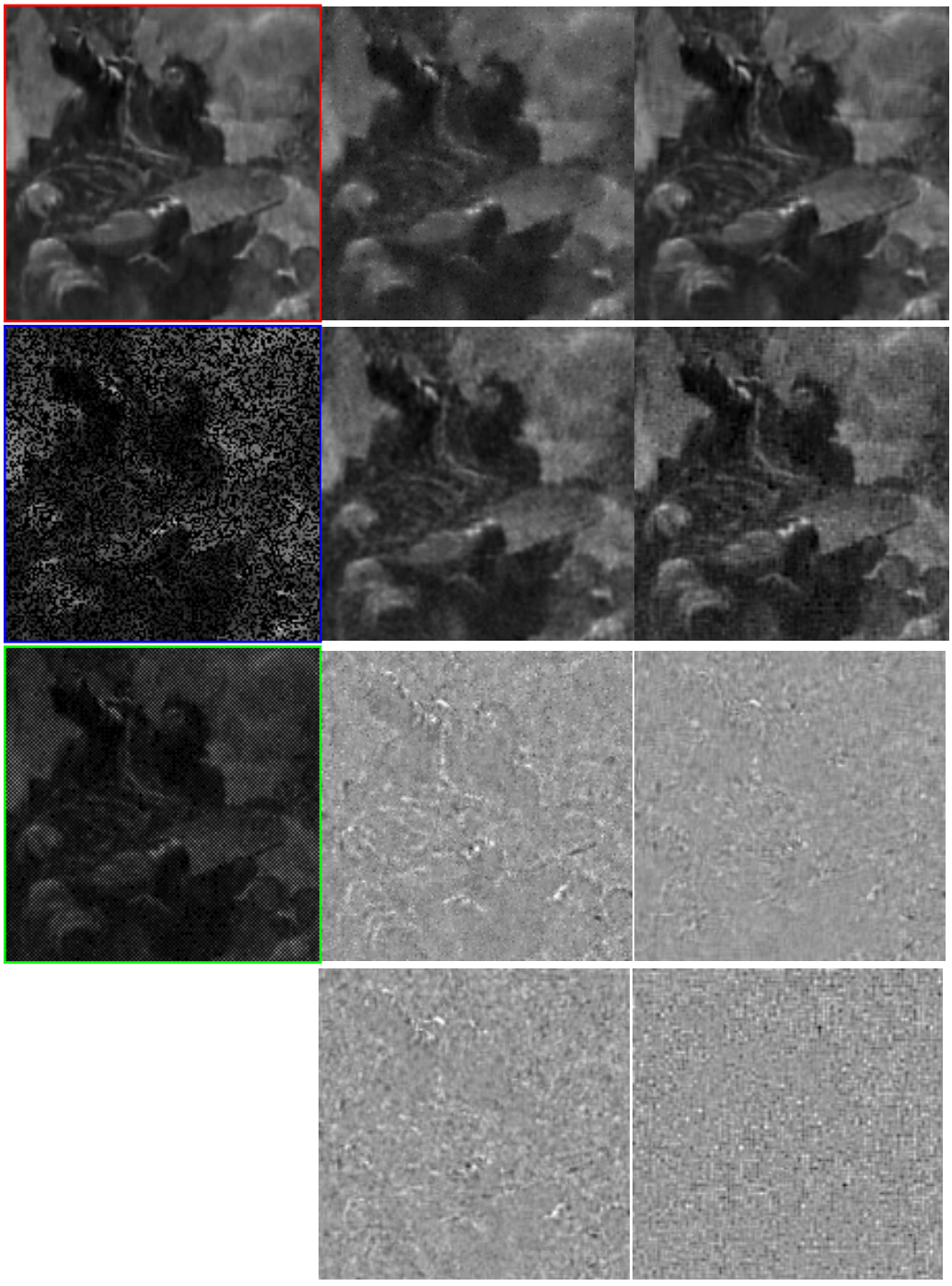


Figure 6.16: **Synthetic data.** 51% missing pixels (for both random and regular pattern). **First column:** Ground-truth (red), input image with random pattern (blue), input image with regular pattern (green). **Second column:** Result by: PLEV, Schöberl et al., and their corresponding difference images with respect to the ground-truth. **Third column:** Result by: proposed method, Nayar and Mitsunaga, and their corresponding difference images with respect to the ground-truth. See PSNR values for these extracts in Table 6.7. Please see the digital copy for better details reproduction.

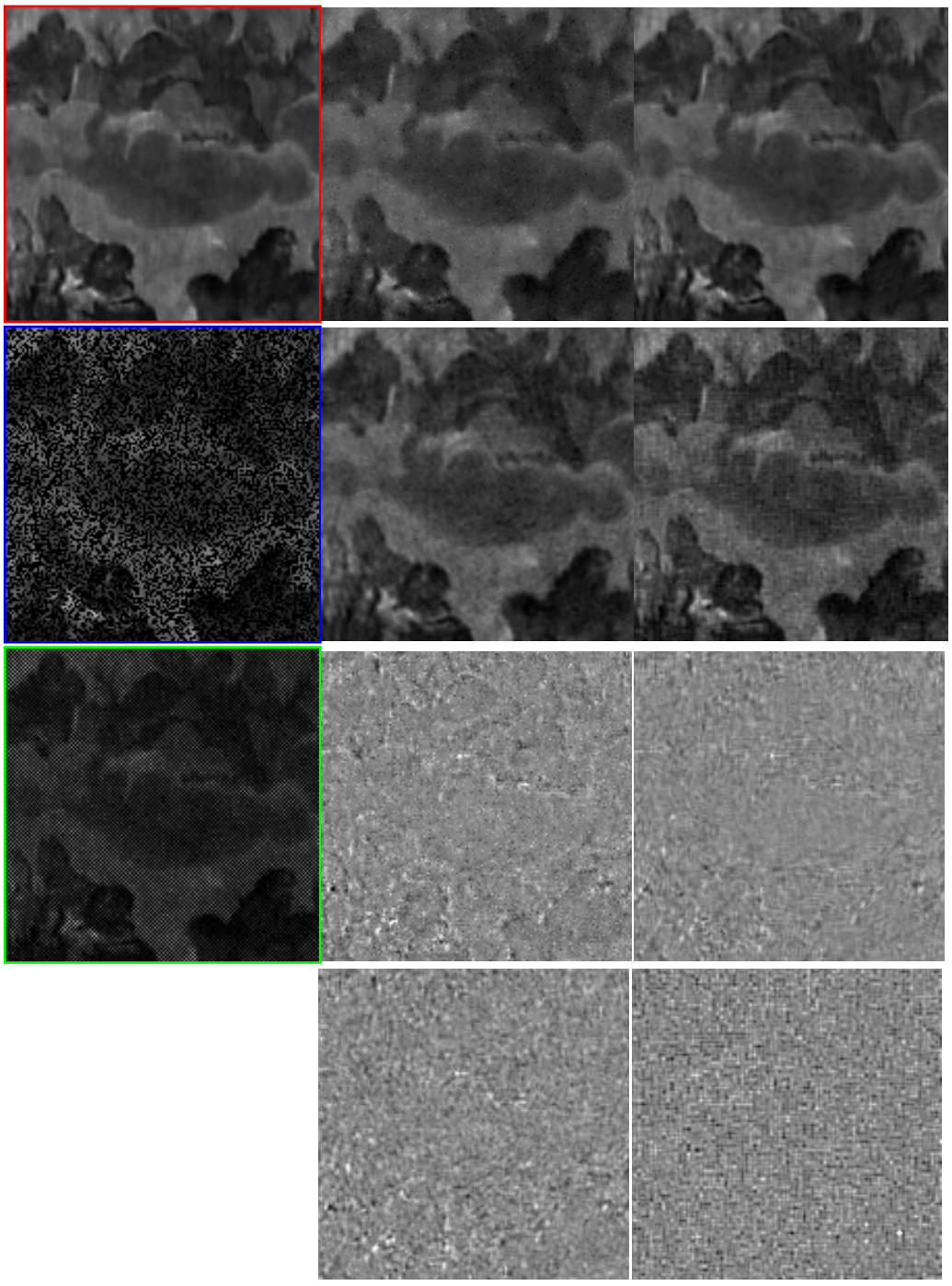


Figure 6.17: **Synthetic data.** 50% missing pixels (for both random and regular pattern). **First column:** Ground-truth (red), input image with random pattern (blue), input image with regular pattern (green). **Second column:** Result by: PLEV, Schöberl et al., and their corresponding difference images with respect to the ground-truth. **Third column:** Result by: proposed method, Nayar and Mitsunaga, and their corresponding difference images with respect to the ground-truth. See PSNR values for these extracts in Table 6.7. Please see the digital copy for better details reproduction.

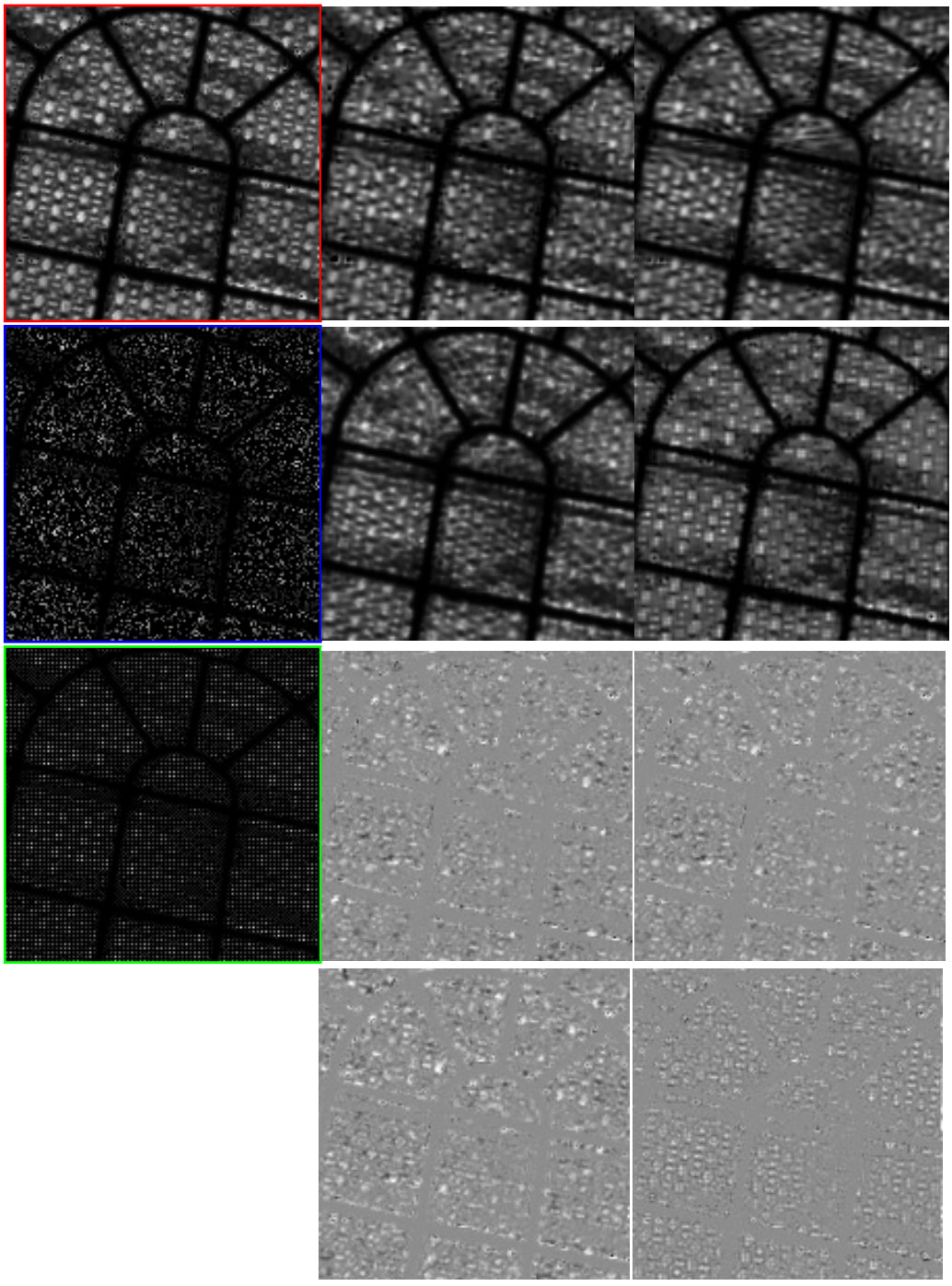


Figure 6.18: **Synthetic data.** 64% missing pixels (for both random and regular pattern). **First column:** Ground-truth (red), input image with random pattern (blue), input image with regular pattern (green). **Second column:** Result by: PLEV, Schöberl et al., and their corresponding difference images with respect to the ground-truth. **Third column:** Result by: proposed method, Nayar and Mitsunaga, and their corresponding difference images with respect to the ground-truth. See PSNR values for these extracts in Table 6.7. Please see the digital copy for better details reproduction.

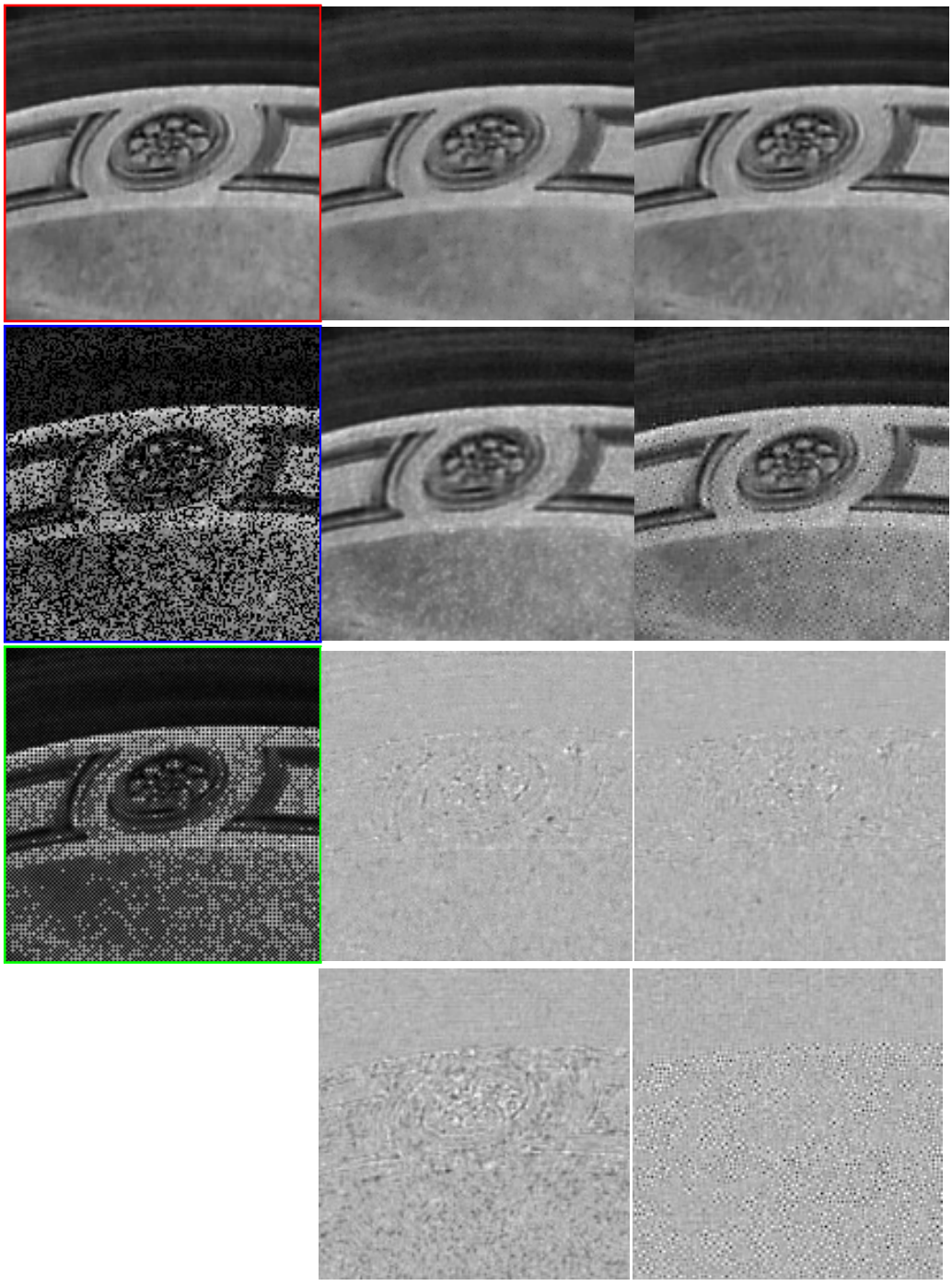


Figure 6.19: **Synthetic data.** 43% missing pixels (for both random and regular pattern). **First column:** Ground-truth (red), input image with random pattern (blue), input image with regular pattern (green). **Second column:** Result by: PLEV, Schöberl et al., and their corresponding difference images with respect to the ground-truth. **Third column:** Result by: proposed method, Nayar and Mitsunaga, and their corresponding difference images with respect to the ground-truth. See PSNR values for these extracts in Table 6.7. Please see the digital copy for better details reproduction.

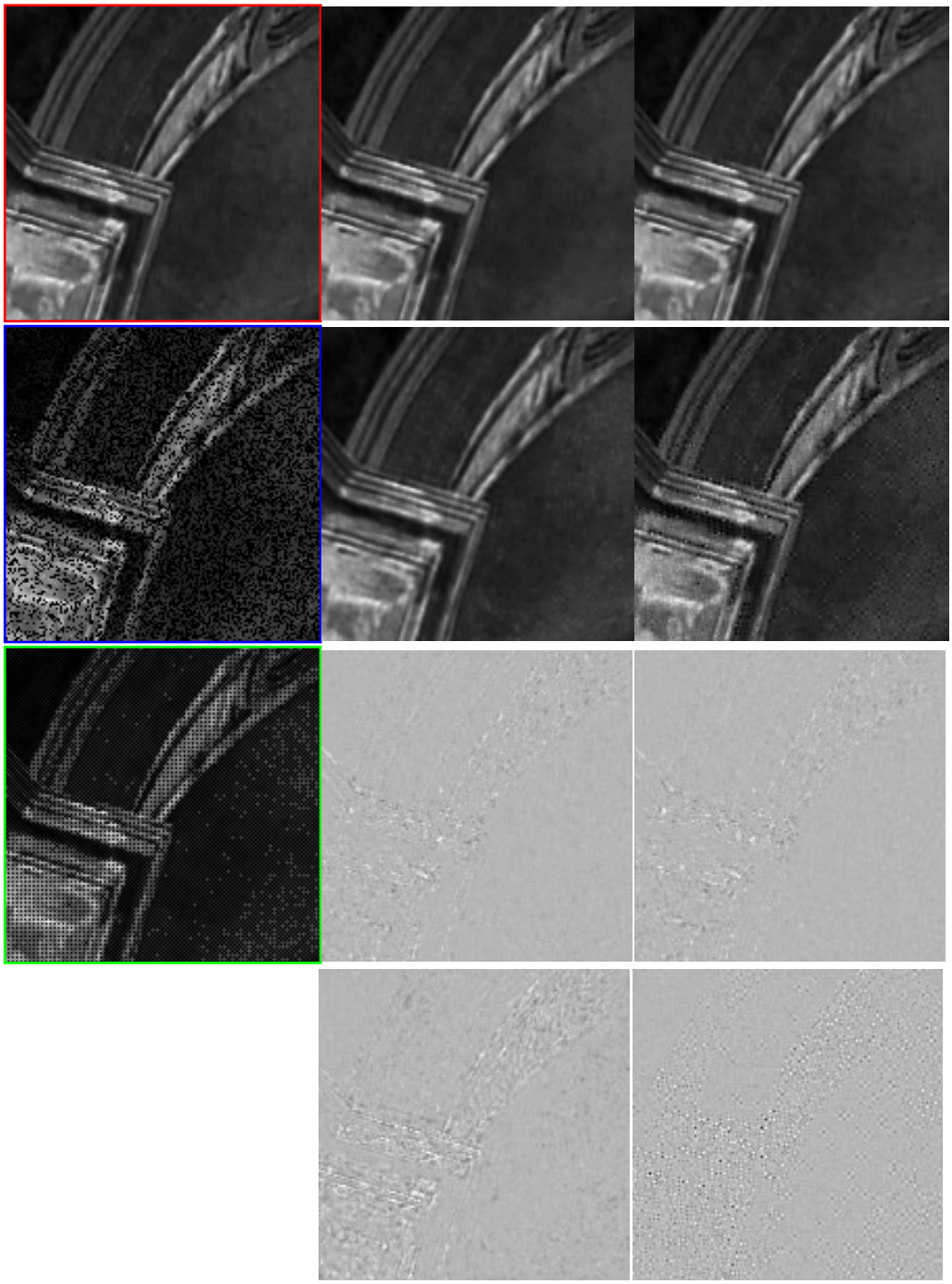


Figure 6.20: **Synthetic data.** 41% missing pixels. Patch size 8×8. **Left to right:** Ground-truth, result obtained by: PLE, proposed approach, Schöberl et al., Nayar and Mitsunaga. Input image and difference images. See PSNR results in Table 6.7. Please see the digital copy for better details reproduction.

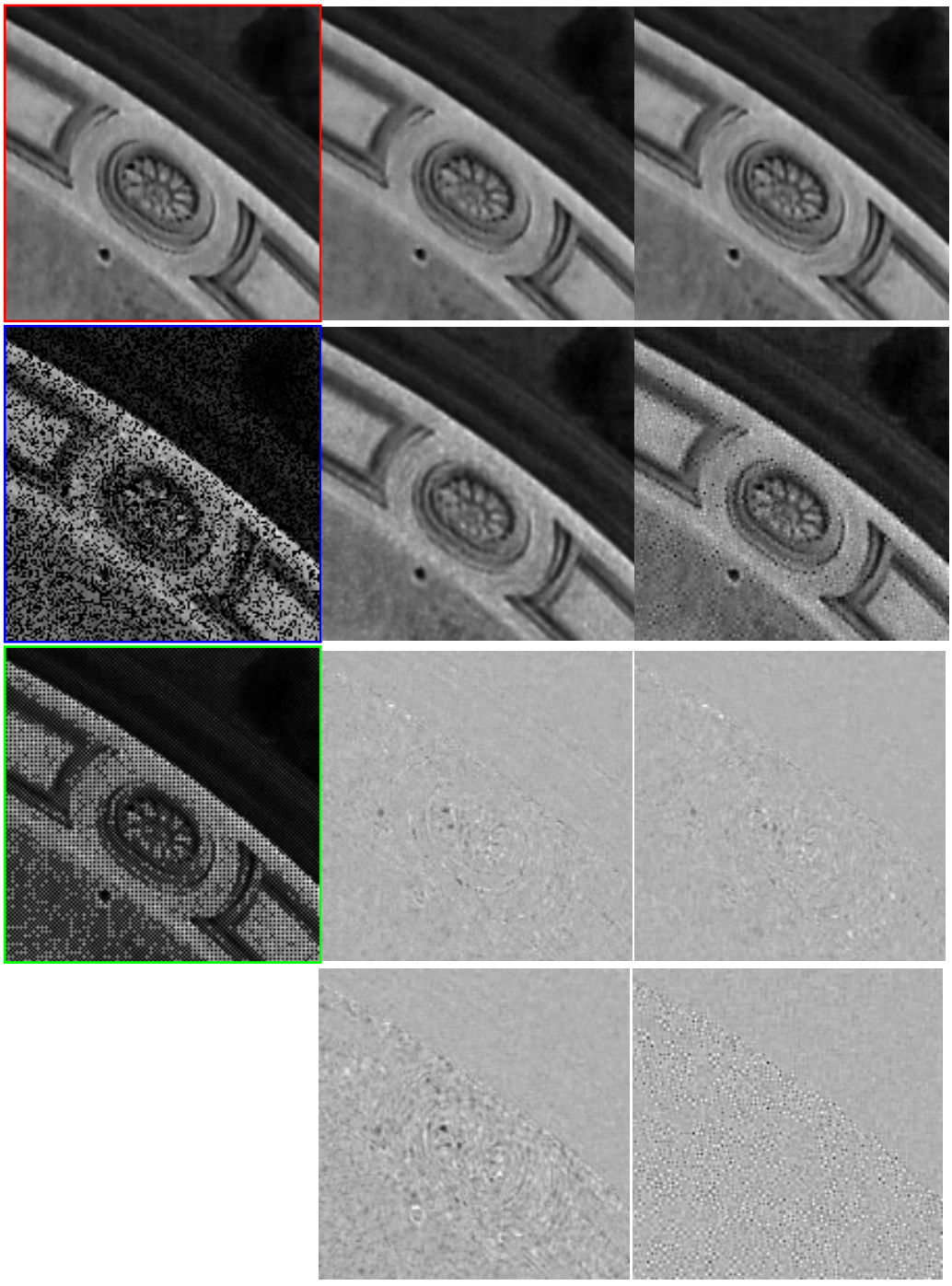


Figure 6.21: **Synthetic data.** 41% missing pixels (for both random and regular pattern). **First column:** Ground-truth (red), input image with random pattern (blue), input image with regular pattern (green). **Second column:** Result by: PLEV, Schöberl et al., and their corresponding difference images with respect to the ground-truth. **Third column:** Result by: proposed method, Nayar and Mitsunaga, and their corresponding difference images with respect to the ground-truth. See PSNR values for these extracts in Table 6.7. Please see the digital copy for better details reproduction.

systems are still not available for general usage. However, as stated in Section 6.4.2, the only variation between the classical and the SVE acquisition is the optical filter, i.e. the amount of light reaching each pixel. Hence, the noise at a pixel p captured using SVE with an optical gain factor o_p and exposure time τ/o_p and a pixel captured with a classical camera using exposure time τ should be very close. We take advantage of this fact in order to evaluate the reconstruction performance of the proposed approach using real data. For this purpose we generate an SVE image drawing pixels at random from four raw images acquired with different exposure times. The four different exposure times simulate the different filters of the SVE optical mask. The images are acquired using a remotely controlled camera and a tripod so as to be perfectly aligned. Otherwise, artifacts may appear from the random sampling of the four images to composite the SVE frame. Notice that the SVE image thus obtained is very similar to the one obtained if such an optical filter was placed adjacent to the sensor. This protocol does not allow us to take scenes with moving objects. Let us emphasize, however, that using a real SVE device, this, as well as the treatment of moving camera, would of course not be an issue.

Given the procedure we use to generate the SVE image form the input raw images, the Bayer pattern of the latter is kept in the generated SVE image. The proposed irradiance reconstruction method is thus applied to the raw SVE image with an overlap of $\sqrt{N} - 2$ between patches (i.e. a shift of two pixels) in order to compare pixels of the corresponding color channels. The parameters for the proposed method are set to: patch size 8×8, $\alpha_H = 1$, $\alpha_L = 0.5$ (α_H and α_L define the values for κ and ν, see Section 6.3.4) in all examples. The parameters for the PLEV method are set to: patch size 8×8, $\epsilon = 0.1$. The demosaicking method by [Hamilton and Adams 1997] is then used to obtain a color image form the reconstructed irradiance. To display the results we use the tone mapping technique by Mantiuk et al. [2008].

A comparison against the methods by Nayar and Mitsunaga and Schöberl et al. is not presented since they do not precise in their works how to treat raw images with a Bayer pattern (how to treat color) and therefore an adaptation of their methods should be made in order to process our data.

Figures 6.22 to 6.27 show the results obtained in three real scenes, together with the mask of known (white) and unknown (black) pixels and some extracts. Recall that among the unknown pixels, some of them correspond to saturated pixels and some of them to under-exposed pixels. The proposed method manages to simultaneously denoise and reconstruct the unknown pixels in quite extreme conditions (more than 70% of the pixels are missing). See the extracts for a comparison between the denoising capacity of PLEV and the proposed methods. These examples show the capacity of the proposed approach to reconstruct the irradiance information in both very dark and bright regions simultaneously. See for instance the example in Figure 6.25, where the dark interior of the building (which can be seen through the windows) and the highly illuminated part of another building are both correctly reconstructed. Figure 6.28 is another example of this where the dark region is particularly dark (please consult the pdf version of this chapter for better visualization).

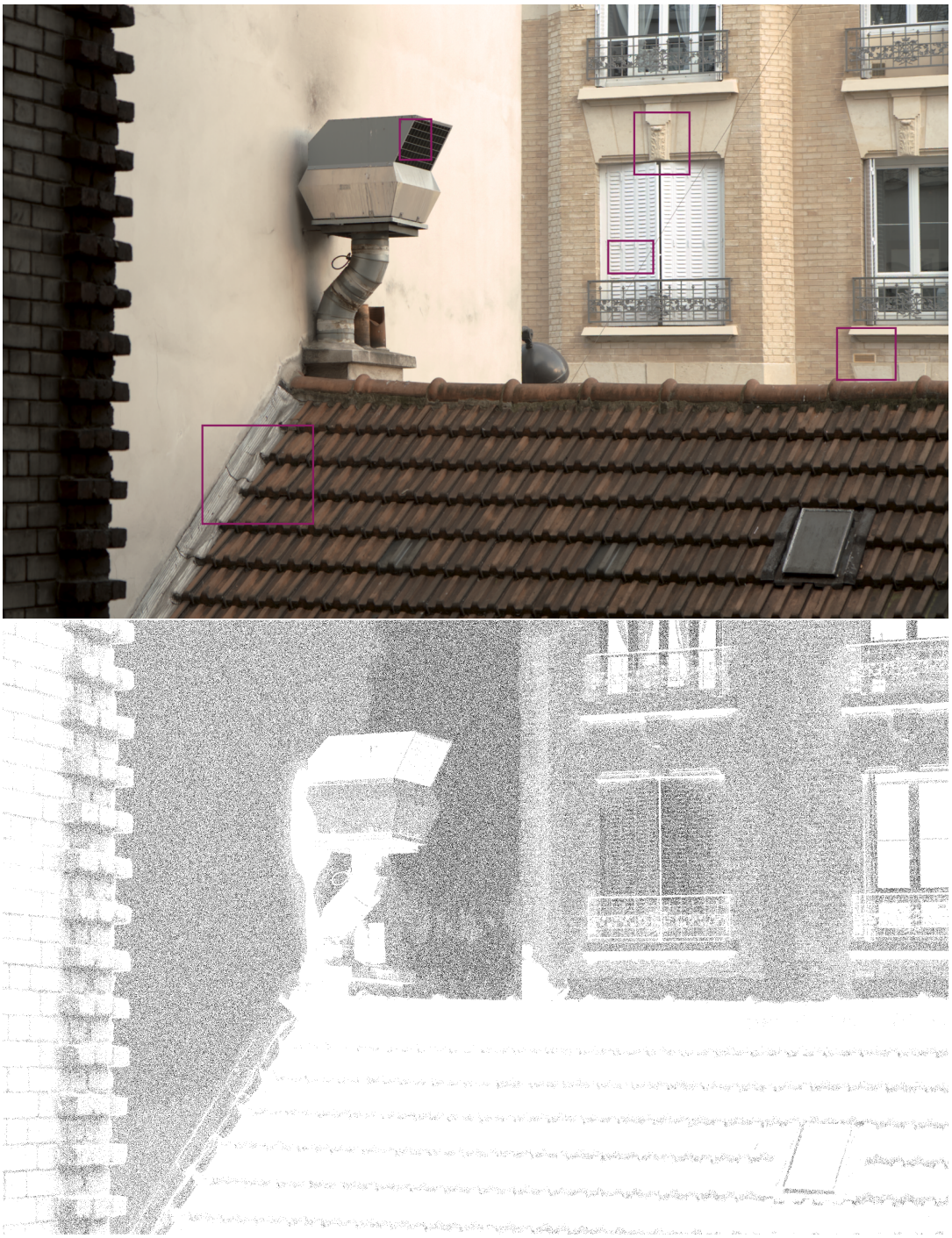


Figure 6.22: **Real data.** Tone mapped version of the HDR image obtained by the proposed approach (11.4 stops) and its corresponding mask of missing pixels (black). Marked regions appear in Figures 6.23 and 6.24. Please see the digital copy for better details reproduction.

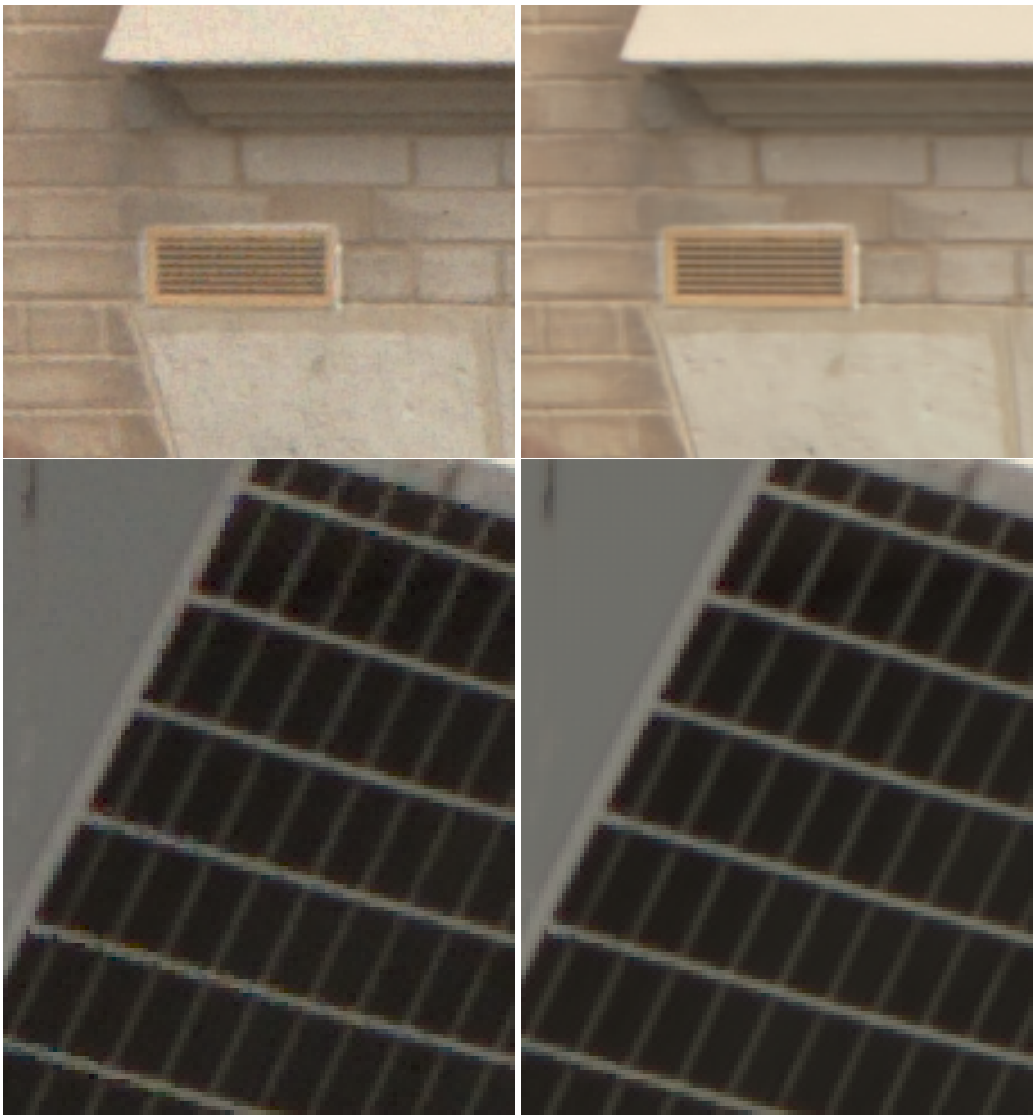


Figure 6.23: **Real data.** Zoom of the extracts in Figure 6.22. **First column:** Results by PLEV. **Second column:** Results by the proposed approach. Please see the digital copy for better details reproduction.

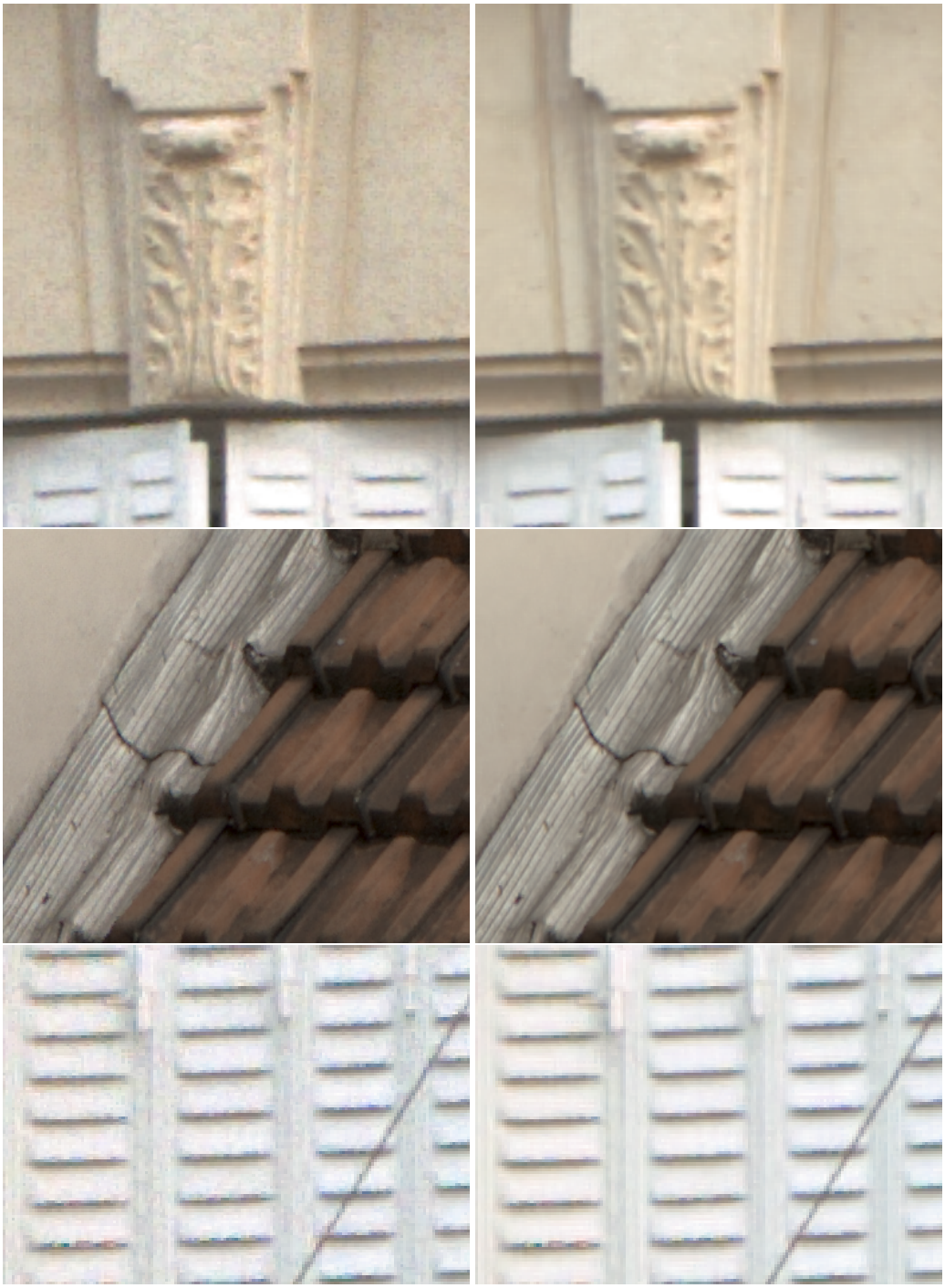


Figure 6.24: **Real data.** Zoom of the extracts in Figure 6.22. **First column:** Results by PLEV. **Second column:** Results by the proposed approach. Please see the digital copy for better details reproduction.

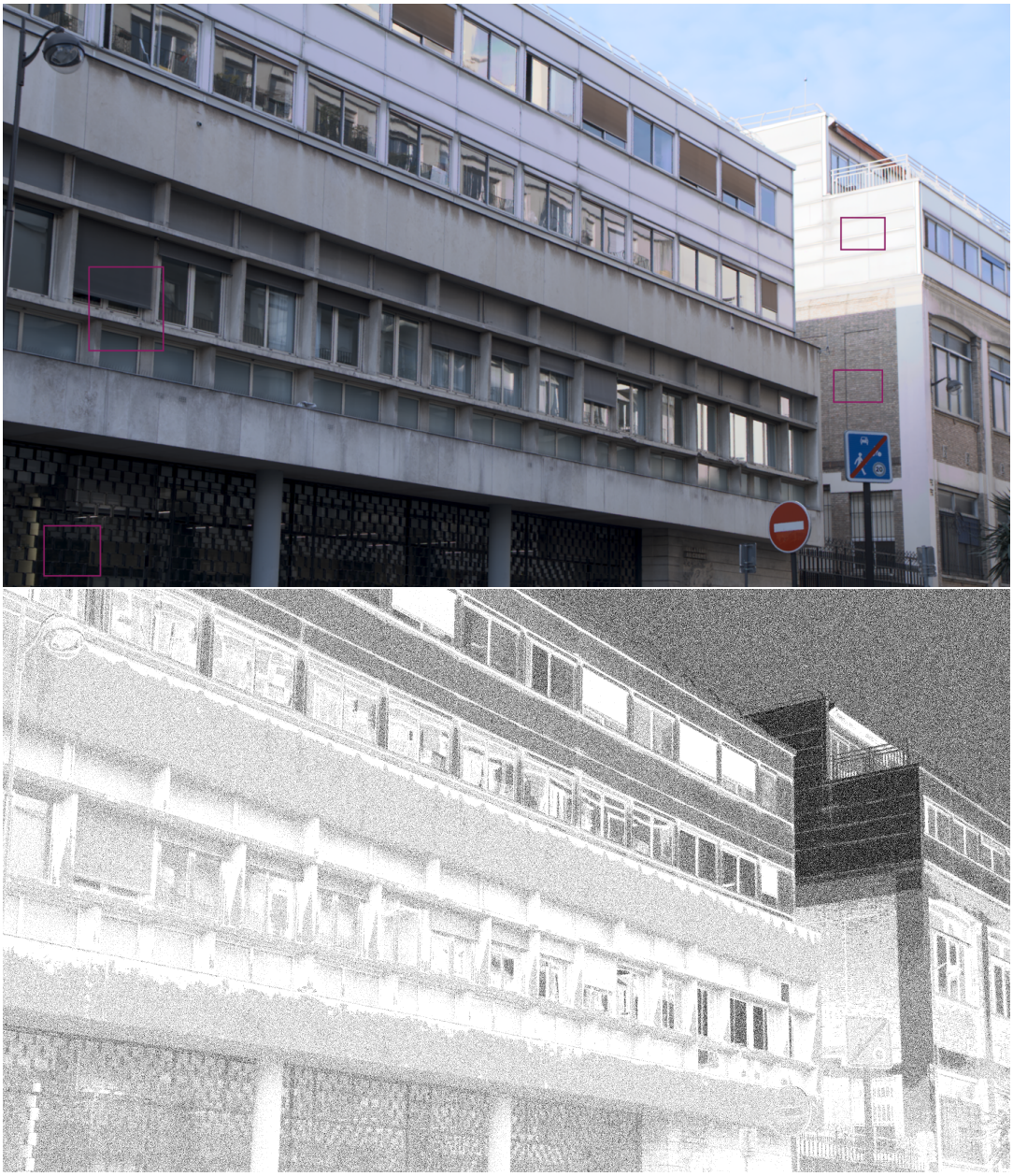


Figure 6.25: **Real data.** Tone mapped version of the HDR image obtained by the proposed approach (15.6 stops) and its corresponding mask of missing pixels (black). Marked regions appear in Figures 6.23 and 6.24. Please see the digital copy for better details reproduction.

Figure 6.26: **Real data.** Zoom of the extracts in Figure 6.25. **First column:** Results by PLEV. **Second column:** Results by the proposed approach. Please see the digital copy for better details reproduction.

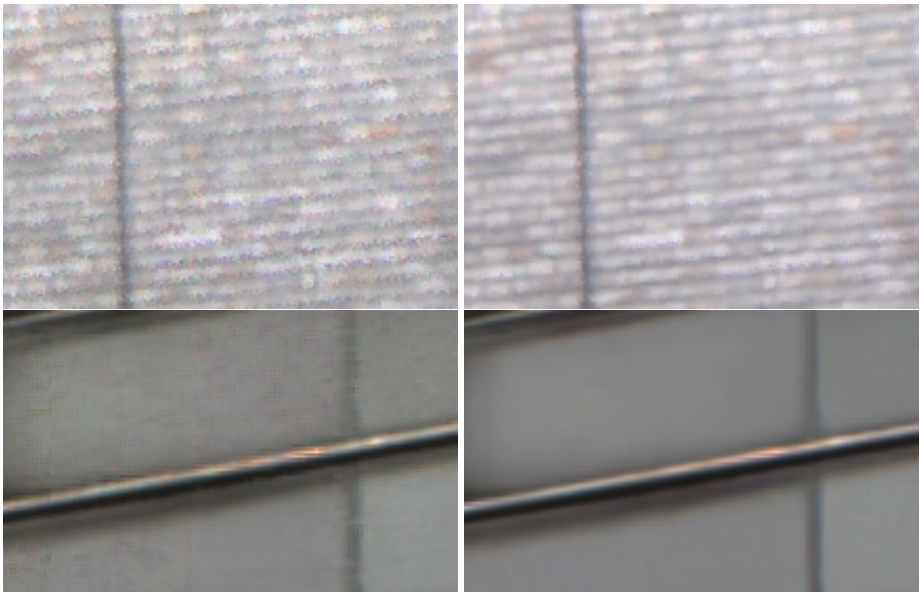


Figure 6.27: **Real data.** Zoom of the extracts in Figure 6.25. **First column:** Results by PLEV. **Second column:** Results by the proposed approach. Please see the digital copy for better details reproduction.

6.5 Conclusions

In this chapter we introduced a new Bayesian patch based reconstruction algorithm that simultaneously handles additive noise and random missing data. The method makes use of an hyperprior on the Gaussian model parameters in order to perform the parameter estimation in the case of missing data. This way, it successfully combines the reconstruction capability of the PLE method with the denoising power of NLB. The reconstruction capability of the proposed method is shown through various experiments conducted with images degraded both by additive noise and random missing pixels. As previously stated, the setting of the hyperprior parameters κ and ν is not a trivial task and should the subject of further study in a future work.

Moreover, we present an application of the proposed method to the generation of HDR images acquired with the single-image SVE strategy. The reconstruction method is clearly adapted to this case since it presents additive noise, with variable variance dependent on the pixel values, and random missing pixels, due to under and over saturated pixels. Experiments conducted with synthetic and real data, acquired in similar conditions to the real SVE acquisition, show the capacity of the method to generate HDR images.

A distance taking into account the variable variance of the pixels, as was considered in the HDR imaging method introduced in Chapter 5, can be also useful in this method since the compared patches have different noise levels.

In the proposed approach, both saturated and under-exposed pixels are equally treated as missing pixels. However, valuable information exists in the fact that a pixel is either saturated or under-exposed [Aguerrebere et al. 2014b](Chapter 3). Future work should explore the possibility of different treatments for each of these two kind of pixels. It would not be surprising that this strategy, if well implemented, may improve current results.

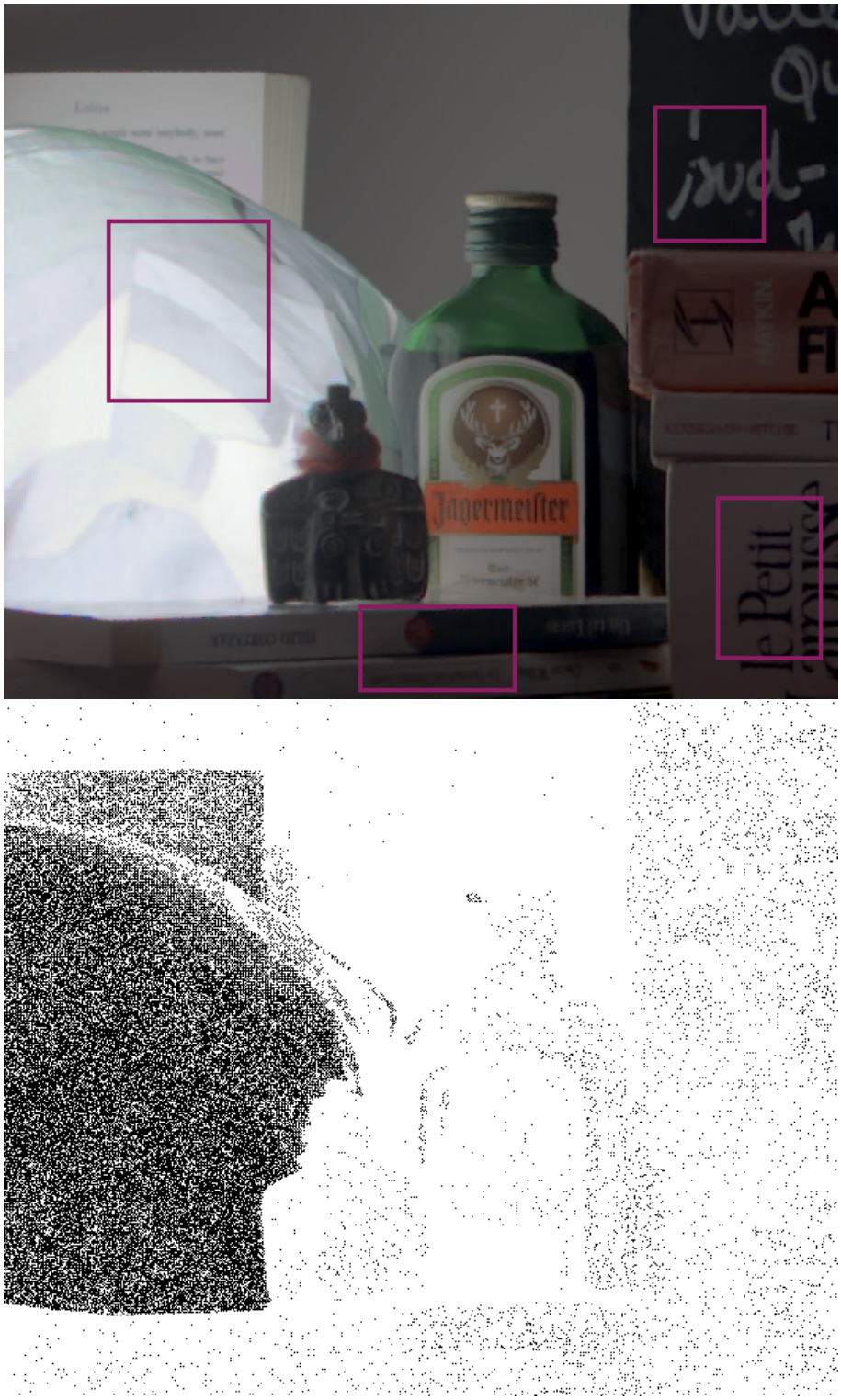


Figure 6.28: **Real data.** Tone mapped version of the HDR image obtained by the proposed approach (13.4 stops) and its corresponding mask of missing pixels (black). Marked regions appear in Figures 6.29 and 6.30. Please see the digital copy for better details reproduction.

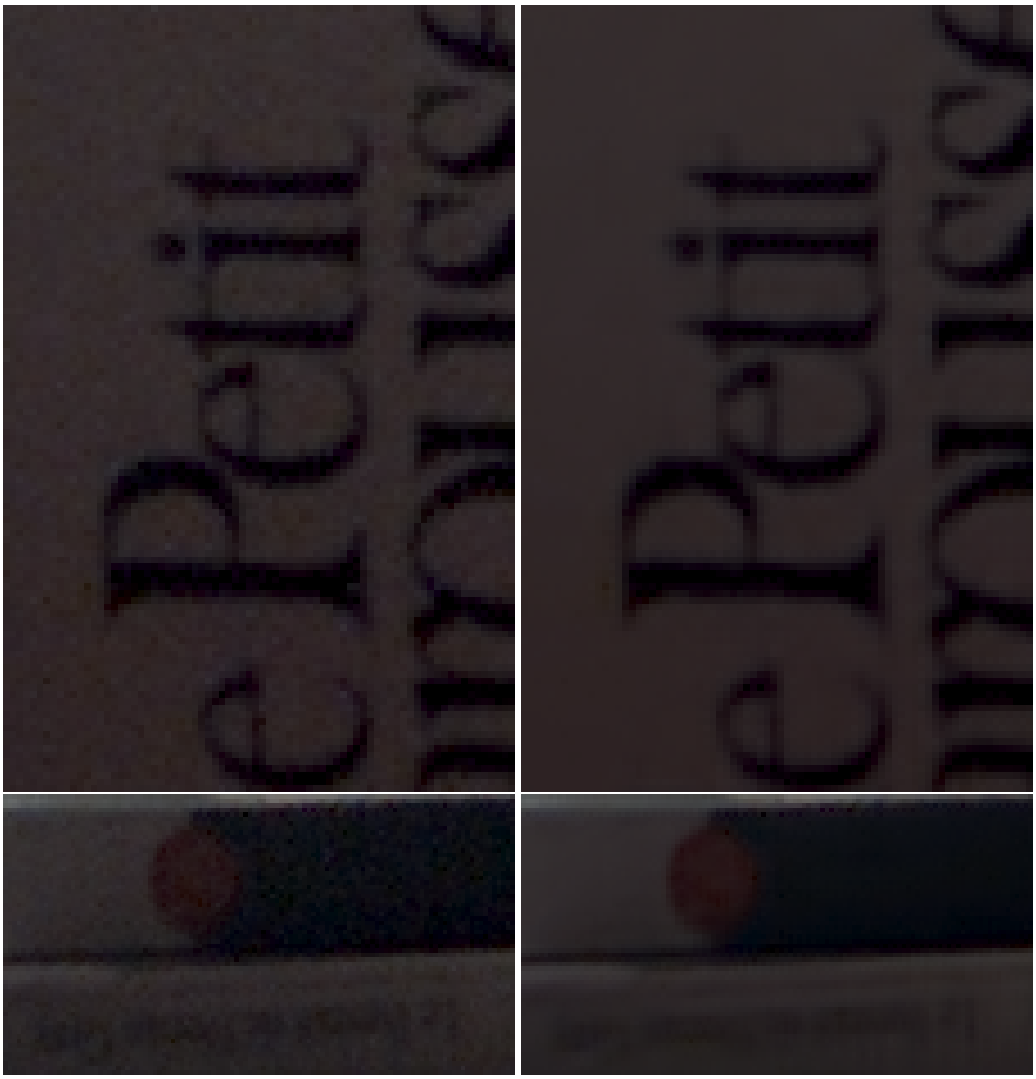


Figure 6.29: **Real data.** Zoom of the extracts in Figure 6.28. **First column:** Results by PLEV. **Second column:** Results by the proposed approach. Please see the digital copy for better details reproduction.

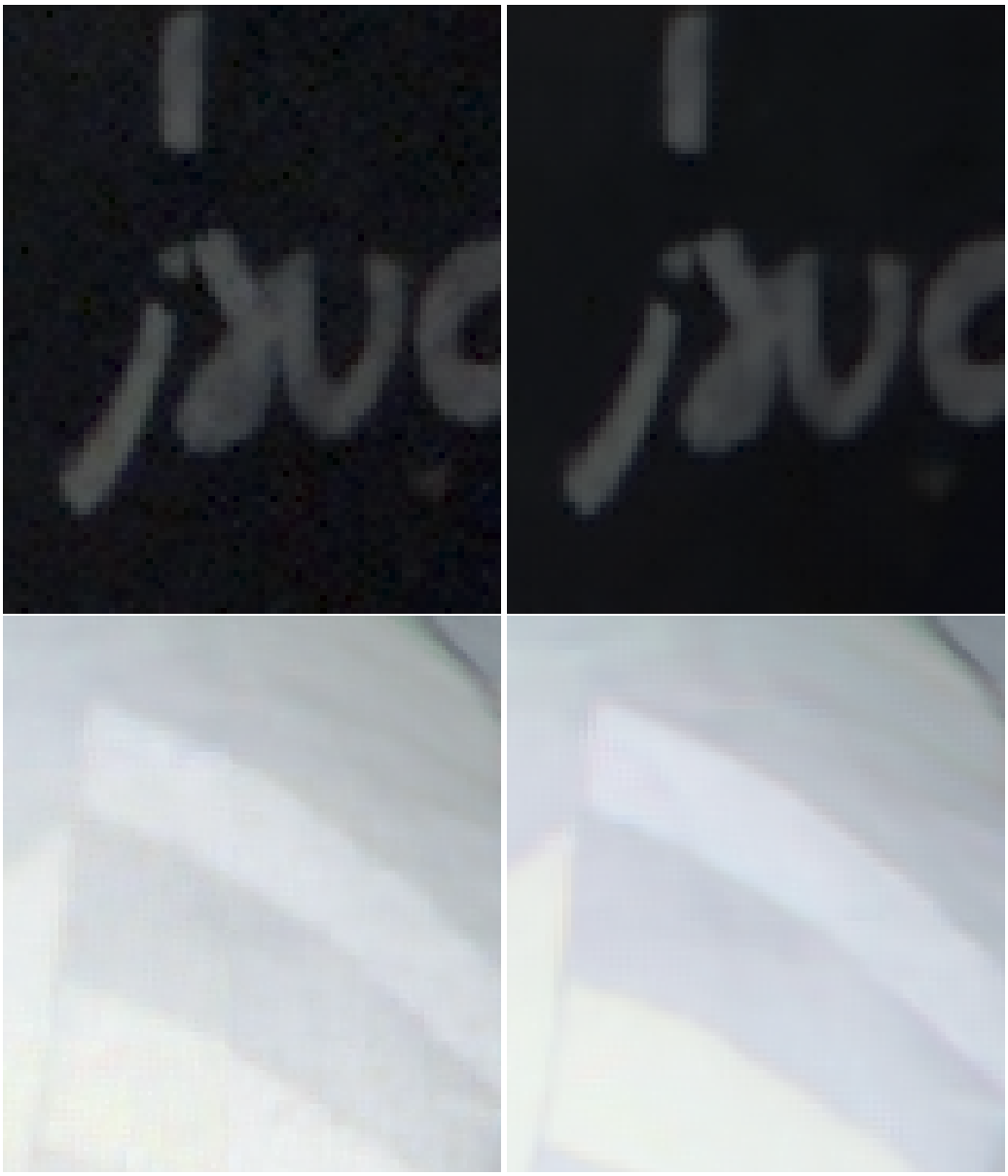


Figure 6.30: **Real data.** Zoom of the extracts in Figure 6.28. **First column:** Results by PLEV. **Second column:** Results by the proposed approach. Please see the digital copy for better details reproduction.

7 Conclusions and Perspectives

In this thesis we presented a thorough study of the generation of HDR images, both from a theoretical point of view and with practical solutions to the problem. The starting point of the work was the study of the camera acquisition process and the proposal of an accurate and realistic, yet simple, noise model. The problem of the generation of HDR images was then stated from a statistical perspective. More precisely, we considered the estimation of the irradiance map from the observations of a set of random variables following the noise model. This problem was analyzed in two different scenarios: the static and the dynamic case.

The performance bounds for the irradiance estimation problem in the static case were computed by means of the Cramér Rao lower bound theorem. We then showed that this bound cannot be attained and therefore no efficient estimator exists for the irradiance under the considered hypotheses: pixel-wise estimation from T independent samples following Model (2.2). The performance of various existing irradiance estimators was compared to the CRLB. We found that all tested estimators perform close to the bound, and in particular, that an iterative approximation of the MLE [Granados et al. 2010] is nearly optimal. This result shows that there is not much room for improvement under the considered hypotheses. Hence, this motivates the search for alternatives to the pixel-wise unbiased estimators classically used in the static case.

One option is the inclusion of the saturated samples. Following the statistical study, a new irradiance estimator was proposed that, unlike existing estimators, includes saturated samples. We showed that the information carried by the fact that a pixel saturates for a given exposure time can be used to improve the estimation of the irradiance values that are close to saturation. This was confirmed by the comparison of the performance bounds computed with and without the saturation information, as well as by numerical experiments.

Another option, taking advantage of the redundancy existing in natural images, is to release the pixel-wise hypothesis and perform the irradiance estimation combining samples from different pixel positions. This idea, inspired from the classical denoising techniques, is exploited in the algorithms developed for the dynamic case in Chapters 5 and 6 and was shown to highly improve the irradiance estimation.

Finally, another interesting option may be the computation of a biased estimator. Biased estimators have been shown useful in various applications, where the bias - variance trade-off can be controlled to improve the obtained results. It is not a priori obvious which

kind of bias would be appropriate for the irradiance estimation and further analysis should be done in this sense.

Keeping the hypothesis of static scene and allowing small hand-held motion for the camera, we studied the combined problem of HDR and super-resolution. From a collaboration with Yann Traonmilin, we proposed an acquisition strategy that, if the affine motion hypothesis holds and sufficiently long exposure time is available, guarantees the recovery of the real high dynamic range and high frequency information of the original scene.

Next, in order to tackle the dynamic case of HDR image generation, where moving objects are present and the acquisition is done using a hand-held camera, we proposed to use a multi-image patch based approach that simultaneously reconstructs and denoises the irradiance map. The use of patches has two main advantages. First, it avoids the need for global image alignment and objects motion detection. This is particularly useful since the ghosting artifacts resulting from image misalignment are extremely annoying. Second, as previously mentioned, the irradiance estimation is performed combining several samples (usually much more than in the static case, i.e., the number of input images) which gives a denoised result.

The main drawback of this approach is the need for filling possibly large saturated areas in the reference frame. We proposed to tackle this inpainting problem using the Poisson editing technique. This is usually possible since the information missing in the reference image can be found in at least one of the other input frames. Nevertheless, other filling approaches should be studied. One possibility, inspired by the texture synthesis technique by Efros and Leung [Efros and Leung 1999], could be using a patch-based method to inpaint the unknown region taking the patches from the input frames where the region is correctly exposed.

An interesting perspective for the improvement of this method is to include the results of the restoration method introduced in Chapter 6, which were not included at first due to the chronological order of development of both methods.

Further, a generic Bayesian patch-based restoration method was proposed that simultaneously handles additive noise and random missing pixels. This method was inspired by the two powerful restoration methods, the Piece-wise Linear Estimators (PLE) [Yu et al. 2012] and the Non local Bayes (NLB) [Lebrun et al. 2013]. As in NLB, a Gaussian model is assumed for image patches, whose parameters are estimated locally from similar patches. The inclusion of an hyperprior on the Gaussian parameters makes it possible to extend the denoising ability of NLB to the case of random missing pixels. Results show that the proposed method keeps the inpainting capability of PLE while boosting its denoising performance through the local model estimation inspired by NLB.

The setting of the hyperprior parameters is not a trivial task. A first setting proposal was done in this work but this point should be the subject of further investigation.

Even if the method has only been tested for denoising and restoration of random missing pixels, it may also be conceived as a more general framework for solving inverse problems (such as interpolation of missing pixels on a regular pattern, zooming, deblurring) as performed in [Yu et al. 2012]. Nevertheless, this extension is not obvious and needs further analysis and testing. In particular, some considerations regarding the initialization of

the method should be accounted for as explained in [Yu et al. 2012].

This reconstruction method was applied to the problem of single image HDR generation using the spatially varying pixel exposure (SVE) acquisition strategy. Being a single-image technique, this approach avoids the ghosting problems caused by image misalignment. However, serious noise problems arise as well as the need to inpaint missing pixels (saturated and under-exposed pixels), which makes it a natural application for the proposed restoration method. Hence, the method is adapted to the precise noise present in these images in order to perform the simultaneous denoising and missing pixel restoration. Examples using synthetic and real data acquired in similar conditions to that of SVE show the ability of the method for the HDR image generation task.

The patch comparison, which is a very important aspect for the methods presented in Chapters 5 and 6, should be further studied. Other approaches such as the likelihood-based ones presented in [Deledalle et al. 2012] should be considered.

The vast experimental testing, done for the evaluation of the different proposed algorithms, confirmed the importance of the choice of the exposure times on the quality of the obtained irradiance map. As Model (2.2) shows, the noise level of the images, and so their signal-to-noise ratio, is partly determined by the exposure time, thus affecting the final result. As presented in Chapter 4, a trade-off must be made between the quality of the correctly exposed pixels and the number of saturated ones. This aspect may be an interesting problem for further analysis.

Finally, let us conclude by observing that an interesting continuation of this work may be the study of the aforementioned problems in the context of the new image sensor technologies developed for HDR imaging in recent years, as for example, the Super CCD EXR sensor developed by Fuji Film. New problems arise with these new technologies, as well as new ways of thinking the HDR imaging.

Appendices

A Tone mapping operators

Tone mapping operators (TMOs) are a fundamental tool in HDR imaging since the dynamic range of these images greatly exceeds the contrast range possibly handled by most conventional displays. These operators allow us to display HDR images so that they look as similar as possible to the real scenes intended to be registered. A wide variety of TMOs can be found in the literature. Here, we present a very brief summary of the existing methods. For a through classification and an in-depth analysis of TMOs we refer the reader to the book by Reinhard et al. [2010].

Many of the existing TMOs are inspired by the behavior of the human visual system (HVS) [Tumblin and Rushmeier 1993; Ward 1994; Ferwerda et al. 1996; Ward et al. 1997; Tumblin et al. 1999; Pattanaik et al. 1998; Ashikhmin 2002; Ferradans et al. 2011]. Making use of visual models, these perception based methods aim at producing a tone mapped image similar to the result obtained by the HVS for the same scene. Among the perception based TMOs, we find the sigmoidal tone reproduction operators based on photoreceptor models. Electrophysiology studies show that the output voltage produced by photoreceptors as a function of light intensity can be accurately described by a sigmoid-like function [Reinhard et al. 2010]. Examples of these methods are the rational quantization function proposed by Schlick [1995], inspired by this there is the work by Tumblin et al. [1999], and the photoreceptor adaptation models by Pattanaik et al. [2000], Pattanaik and Yee [2002] and Reinhard and Devlin [2005].

Another example of sigmoide-like tone reproduction operator is the photographic tone reproduction. The idea is to inspire the tone mapping operators in the techniques used in conventional photography developed to tackle the HDR imaging problem. One such approach, proposed by Reinhard et al. [2002] consists in a initial linear scaling of the image followed by a contrast adjustement, that can be global or local, inspired in the dodging-and-burning technique used in conventional photography. This technique consists in exposing a print for longer or shorter periods so as to control the exposure level of different image regions according to its content (darker or brighter). Other examples of HVS-based models include brightness preserving operators [Tumblin and Turk 1999] and retinex-based operators [Rahman et al. 1996; Jobson et al. 1997].

Another class of TMOs tackles the problem from a different perspective, not necessarily driven by visually inspired models but by mathematical or engineering principles. Among these we find the histogram adjustment technique developed by Ward et al. [1997], the fast bilateral filtering implementation by Durand and Dorsey [2002] and the gradient

domain compression by Fattal et al. [2002].

Another possible classification for TMOs is into global and local methods [Devlin et al. 2002]. Global methods apply the same tone mapping operation to all image pixels [Tumblin and Rushmeier 1993; Ward 1994; Pattanaik et al. 2000; Ferwerda et al. 1996; Reinhard and Devlin 2005]. The contrast adjustment parameters are often computed from the image extrema and the mean irradiance of the whole image. On the one hand, the global methods have the advantage of being computationally efficient and not presenting artifacts. On the other hand, they may have limited performance in examples of very large dynamic range, since the globally chosen parameters might not be accurate enough to represent regions of extreme luminance. For this reason, local methods are usually preferred, since they can outperform global methods in local contrast enhancement. Local TMOs apply a different contrast transformation to each pixel [Durand and Dorsey 2002; Reinhard et al. 2002; Fattal et al. 2002; Ferradans et al. 2011], this transformation being often defined by the luminance level of a set of neighboring pixels. A common drawback of local TMOs are the halo artifacts, caused by an incorrect setting of the transformation parameters due to large lighting differences between the affected pixel and its neighborhood. Combined spatial and radiometric distances, such as in the bilateral filtering [Durand and Dorsey 2002], are of great help to reduce this kind of artifact.

B MAP estimation of parameters

Let us consider the log-likelihood

$$\ln p(\mu, \mathbf{\Lambda}|\mathrm{z}_1, \ldots, \mathrm{z}_N) = \ln p(\mathrm{z}_1, \ldots, \mathrm{z}_N|\mu, \Sigma) + \ln p(\mu, \mathbf{\Lambda}) \tag{B.1}$$

$$\propto \frac{1}{2}\sum_{j=1}^{N} \ln|\mathbf{\Lambda}_j^*| - \frac{1}{2}\sum_{j=1}^{N}(\mathrm{z}_j - \boldsymbol{D}_j\mu)\mathbf{\Lambda}_j^*(\mathrm{z}_j - \boldsymbol{D}_j\mu)^T \tag{B.2}$$

$$- \frac{\kappa}{2}(\mu - \mu_0)\mathbf{\Lambda}(\mu - \mu_0)^T + \frac{(\nu - d)}{2}\ln|\mathbf{\Lambda}| - \frac{1}{2}\mathrm{tr}(\nu\Sigma_0\mathbf{\Lambda}), \tag{B.3}$$

with $\mathbf{\Lambda}_j^* = (\boldsymbol{D}_j\mathbf{\Lambda}^{-1}\boldsymbol{D}_j^T + \Sigma_w)^{-1}$.

The partial derivative of $\ln p(\mu, \mathbf{\Lambda}|\mathrm{z}_1, \ldots, \mathrm{z}_N)$ with respect to μ is

$$\frac{\partial p(\mu, \mathbf{\Lambda}|\mathrm{z}_1, \ldots, \mathrm{z}_N)}{\partial \mu} = \sum_{j=1}^{N} \boldsymbol{D}_i\mathbf{\Lambda}_j^*(\mathrm{z}_j - \boldsymbol{D}_j\mu) - \kappa\mathbf{\Lambda}(\mu - \mu_0), \tag{B.4}$$

Equating to zero we have

$$\sum_{j=1}^{N} \boldsymbol{D}_i\mathbf{\Lambda}_j^*(\mathrm{z}_j - \boldsymbol{D}_j\mu) - \kappa\mathbf{\Lambda}(\mu - \mu_0) = 0 \tag{B.5}$$

$$\sum_{j=1}^{N} \boldsymbol{D}_i\mathbf{\Lambda}_j^*\mathrm{z}_j - \sum_{j=1}^{N} \boldsymbol{D}_i\mathbf{\Lambda}_j^*\boldsymbol{D}_j\mu - \kappa\mathbf{\Lambda}(\mu - \mu_0) = 0 \tag{B.6}$$

$$\mathbf{\Lambda}\sum_{j=1}^{N} \mathbf{\Lambda}^{-1}\boldsymbol{D}_i\mathbf{\Lambda}_j^*\mathrm{z}_j - \mathbf{\Lambda}\sum_{j=1}^{N} \mathbf{\Lambda}^{-1}\boldsymbol{D}_i\mathbf{\Lambda}_j^*\boldsymbol{D}_j\mu - \kappa\mathbf{\Lambda}(\mu - \mu_0) = 0 \tag{B.7}$$

$$\left(\kappa Id + \sum_{j=1}^{N} \mathbf{\Lambda}^{-1}\boldsymbol{D}_i\mathbf{\Lambda}_j^*\boldsymbol{D}_j\right)\mu - \sum_{j=1}^{N} \mathbf{\Lambda}^{-1}\boldsymbol{D}_i\mathbf{\Lambda}_j^*\mathrm{z}_j + \kappa\mu_0 = 0. \tag{B.8}$$

Thus

$$\mu = \left(\kappa Id + \sum_{j=1}^{N} \mathbf{\Lambda}^{-1}\boldsymbol{D}_i\mathbf{\Lambda}_j^*\boldsymbol{D}_j\right)^{-1}\left(\sum_{j=1}^{N} \mathbf{\Lambda}^{-1}\boldsymbol{D}_i\mathbf{\Lambda}_j^*\mathrm{z}_j + \kappa\mu_0\right). \tag{B.9}$$

The partial derivative of $\ln p(\mu, \mathbf{\Lambda}|\mathrm{z}_1, \ldots, \mathrm{z}_N)$ with respect to $\mathbf{\Lambda}$ is

$$\frac{\partial p(\mu, \mathbf{\Lambda}|\mathrm{z}_1, \ldots, \mathrm{z}_N)}{\partial \mathbf{\Lambda}} = \frac{1}{2}\sum_{j=1}^{N}\frac{\partial \ln|\mathbf{\Lambda}_j^*|}{\partial \mathbf{\Lambda}} - \frac{1}{2}\sum_{j=1}^{N}\frac{\partial(\mathrm{z}_j - \boldsymbol{D}_j\mu)\mathbf{\Lambda}_j^*(\mathrm{z}_j - \boldsymbol{D}_j\mu)^T}{\partial \mathbf{\Lambda}} \tag{B.10}$$

$$- \frac{\kappa}{2}\frac{\partial(\mu - \mu_0)\mathbf{\Lambda}(\mu - \mu_0)^T}{\partial \mathbf{\Lambda}} + \frac{(\nu - d)}{2}\frac{\partial \ln|\mathbf{\Lambda}|}{\partial \mathbf{\Lambda}} \tag{B.11}$$

$$- \frac{1}{2}\frac{\partial \mathrm{tr}(\nu\Sigma_0\mathbf{\Lambda})}{\partial \mathbf{\Lambda}}. \tag{B.12}$$

And, assuming that $\mathbf{\Lambda}$ is a symmetric matrix, we have,

$$\frac{\partial \ln|\mathbf{\Lambda}_j^*|}{\partial \mathbf{\Lambda}} = 2(\boldsymbol{D}\mathbf{\Lambda}^{-1})^T\mathbf{\Lambda}^*\boldsymbol{D}\mathbf{\Lambda}^{-1} - ((\boldsymbol{D}\mathbf{\Lambda}^{-1})^T\mathbf{\Lambda}^*\boldsymbol{D}\mathbf{\Lambda}^{-1} \circ Id). \tag{B.13}$$

where $\circ$ is the Hadamard (elementwise) product.

$$\frac{\partial(\mathrm{z}_j - \boldsymbol{D}_j\mu)\mathbf{\Lambda}_j^*(\mathrm{z}_j - \boldsymbol{D}_j\mu)^T}{\partial \mathbf{\Lambda}} = 2\mathbf{\Lambda}^{-1}\boldsymbol{D}^T\mathbf{\Lambda}^*(\mathrm{z}_j - \boldsymbol{D}_j\mu)(\mathrm{z}_j - \boldsymbol{D}_j\mu)^T\mathbf{\Lambda}^*\boldsymbol{D}\mathbf{\Lambda}^{-1} \tag{B.14}$$

$$- (\mathbf{\Lambda}^{-1}\boldsymbol{D}^T\mathbf{\Lambda}^*(\mathrm{z}_j - \boldsymbol{D}_j\mu)(\mathrm{z}_j - \boldsymbol{D}_j\mu)^T\mathbf{\Lambda}^*\boldsymbol{D}\mathbf{\Lambda}^{-1} \circ Id). \tag{B.15}$$

$$\frac{\partial(\mu - \mu_0)\mathbf{\Lambda}(\mu - \mu_0)^T}{\partial \mathbf{\Lambda}} = 2(\mu - \mu_0)(\mu - \mu_0)^T - ((\mu - \mu_0)(\mu - \mu_0)^T \circ Id) \tag{B.16}$$

$$\frac{\partial \ln|\mathbf{\Lambda}|}{\partial \mathbf{\Lambda}} = 2\mathbf{\Lambda}^{-1} - (\mathbf{\Lambda}^{-1} \circ Id) \tag{B.17}$$

$$\frac{\partial \mathrm{tr}(\nu\Sigma_0\mathbf{\Lambda})}{\partial \mathbf{\Lambda}} = \nu\Sigma_0 + \nu\Sigma_0^T - (\nu\Sigma_0 \circ Id) \tag{B.18}$$

$$= 2\nu\Sigma_0 - (\nu\Sigma_0 \circ Id). \tag{B.19}$$

From (B.13), (B.15), (B.16), (B.17), (B.19)

$$\frac{\partial p(\mu, \mathbf{\Lambda}|\mathrm{z}_1, \ldots, \mathrm{z}_N)}{\partial \mathbf{\Lambda}} = \frac{1}{2}\sum_{j=1}^{N} 2(\boldsymbol{D}\mathbf{\Lambda}^{-1})^T\mathbf{\Lambda}^*\boldsymbol{D}\mathbf{\Lambda}^{-1} - ((\boldsymbol{D}\mathbf{\Lambda}^{-1})^T\mathbf{\Lambda}^*\boldsymbol{D}\mathbf{\Lambda}^{-1} \circ Id) \tag{B.20}$$

$$- \frac{1}{2}\sum_{j=1}^{N} 2\mathbf{\Lambda}^{-1}\boldsymbol{D}^T\mathbf{\Lambda}^*(\mathrm{z}_j - \boldsymbol{D}_j\mu)(\mathrm{z}_j - \boldsymbol{D}_j\mu)^T\mathbf{\Lambda}^*\boldsymbol{D}\mathbf{\Lambda}^{-1} \tag{B.21}$$

$$- (\mathbf{\Lambda}^{-1}\boldsymbol{D}^T\mathbf{\Lambda}^*(\mathrm{z}_j - \boldsymbol{D}_j\mu)(\mathrm{z}_j - \boldsymbol{D}_j\mu)^T\mathbf{\Lambda}^*\boldsymbol{D}\mathbf{\Lambda}^{-1} \circ Id) \tag{B.22}$$

$$- \frac{\kappa}{2}(2(\mu - \mu_0)(\mu - \mu_0)^T - ((\mu - \mu_0)(\mu - \mu_0)^T \circ Id)) \tag{B.23}$$

$$+ \frac{(\nu - d)}{2}(2\mathbf{\Lambda}^{-1} - (\mathbf{\Lambda}^{-1} \circ Id)) \tag{B.24}$$

$$- \frac{1}{2}(2\nu\Sigma_0 - (\nu\Sigma_0 \circ Id)). \tag{B.25}$$

Equating to zero we have

$$\sum_{j=1}^{N}(\boldsymbol{D}\boldsymbol{\Lambda}^{-1})^{T}\boldsymbol{\Lambda}^{*}\boldsymbol{D}\boldsymbol{\Lambda}^{-1}-\sum_{j=1}^{N}\boldsymbol{\Lambda}^{-1}\boldsymbol{D}^{T}\boldsymbol{\Lambda}^{*}(\mathrm{z}_{j}-\boldsymbol{D}_{j}\mu)(\mathrm{z}_{j}-\boldsymbol{D}_{j}\mu)^{T}\boldsymbol{\Lambda}^{*}\boldsymbol{D}\boldsymbol{\Lambda}^{-1} \tag{B.26}$$

$$-\kappa(\mu-\mu_{0})(\mu-\mu_{0})^{T}+(\nu-d)\boldsymbol{\Lambda}^{-1}-\nu\Sigma_{0}=0. \tag{B.27}$$

Thus

$$\left((\nu-d)Id+\sum_{j=1}^{N}\boldsymbol{\Lambda}^{-1}\boldsymbol{D}_{i}\boldsymbol{\Lambda}_{j}^{*}\boldsymbol{D}_{j}\right)\boldsymbol{\Lambda}^{-1}=\sum_{j=1}^{N}(\boldsymbol{\Lambda}^{-1}\boldsymbol{D}^{T}\boldsymbol{\Lambda}^{*}(\mathrm{z}_{j}-\boldsymbol{D}_{j}\mu)) \tag{B.28}$$

$$(\boldsymbol{\Lambda}^{-1}\boldsymbol{D}^{T}\boldsymbol{\Lambda}^{*}(\mathrm{z}_{j}-\boldsymbol{D}_{j}\mu))^{T}+\nu\Sigma_{0}. \tag{B.29}$$

and

$$\boldsymbol{\Lambda}^{-1}=\left((\nu-d)Id+\sum_{j=1}^{N}\boldsymbol{\Lambda}^{-1}\boldsymbol{D}_{i}\boldsymbol{\Lambda}_{j}^{*}\boldsymbol{D}_{j}\right)^{-1} \tag{B.30}$$

$$\left(\sum_{j=1}^{N}(\boldsymbol{\Lambda}^{-1}\boldsymbol{D}^{T}\boldsymbol{\Lambda}^{*}(\mathrm{z}_{j}-\boldsymbol{D}_{j}\mu))(\boldsymbol{\Lambda}^{-1}\boldsymbol{D}^{T}\boldsymbol{\Lambda}^{*}(\mathrm{z}_{j}-\boldsymbol{D}_{j}\mu))^{T}+\nu\Sigma_{0}\right). \tag{B.31}$$

Bibliography

Aguerrebere, C., Delon, J., Gousseau, Y., and Musé, P. Study of the digital camera acquisition process and statistical modeling of the sensor raw data. *Preprint HAL http://hal.archives-ouvertes.fr/docs/00/73/35/38/PDF/camera_model.pdf*, 2012.

Aguerrebere, C., Delon, J., Gousseau, Y., and Muse, P. Simultaneous HDR image reconstruction and denoising for dynamic scenes. In *Proceedings of IEEE International Conference on Computational Photography (ICCP)*, pages 1–11, 2013. doi: 10.1109/ICCPhot.2013.6528309.

Aguerrebere, C., Almansa, A., Delon, J., Gousseau, Y., and Muse, P. Single shot high dynamic range imaging using piecewise linear estimators. In *Proceedings of IEEE International Conference on Computational Photography (ICCP)*, 2014a.

Aguerrebere, C., Delon, J., Gousseau, Y., and Musé, P. Best Algorithms for HDR Image Generation. A Study of Performance Bounds. *SIAM Journal on Imaging Sciences*, 7(1):1–34, 2014b. doi: 10.1137/120891952.

Ahuja, N. A. and Bose, N. K. Multidimensional Generalized Sampling Theorem for wavelet Based Image Superresolution. In *Proceedings of IEEE International Conference on Image Processing*, pages 1589–1592, 2006. doi: 10.1109/ICIP.2006.312612.

Ajdari Rad, A., Meylan, L., Vandewalle, P., and Süsstrunk, S. Multidimensional image enhancement from a set of unregistered differently exposed images. In *Proceedings of IS&T/SPIE Electronic Imaging: Computational Imaging V*, volume 6498, 2007. doi: 10.1117/12.704004.

Ashikhmin, M. A tone mapping algorithm for high contrast images. In *Proceedings of the 13th Eurographics Workshop on Rendering*, pages 145–156, 2002. ISBN 1-58113-534-3.

Baker, S. and Kanade, T. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1167–1183, 2002. doi: 10.1109/TPAMI.2002.1033210.

Barakat, N., Hone, A. N., and Darcie, T. E. Minimal-bracketing sets for high-dynamic-range image capture. *IEEE Transactions on Image Processing*, 17(10):1864–1875, 2008. doi: 10.1109/TIP.2008.2001414.

Bengtsson, T., Gu, I. Y., Viberg, M., and Lindström, K. Regularized optimization for joint super-resolution and high dynamic range image reconstruction in a perceptually uniform domain. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1097–1100, 2012. doi: 10.1109/icassp.2012.6288078.

Bogoni, L. Extending dynamic range of monochrome and color images through fusion. In *Proceedings of IEEE International Conference on Pattern Recognition*, volume 3, pages 7–12, 2000. doi: dx.doi.org/10.1109/ICPR.2000.903475.

Boracchi, G. and Foi, A. Multiframe raw-data denoising based on block-matching and 3-D filtering for low-light imaging and stabilization. In *Proceedings of the International Workshop on Local and Non-Local Approximation in Image Processing*, volume 1, 2008.

Brouk, I., Nemirovsky, A., and Nemirovsky, Y. Analysis of noise in CMOS image sensor. In *Proceedings of IEEE International Conference on Microwaves, Communications, Antennas and Electronic Systems (COMCAS)*, pages 1–8, 2008. doi: dx.doi.org/10.1109/COMCAS.2008.4562800.

Buades, A., Coll, B., and Morel, J. M. A Review of Image Denoising Algorithms, with a New One. *SIAM Multiscale Modeling & Simulation*, 4(2):490–530, 2005a. doi: dx.doi.org/10.1137/040616024.

Buades, A., Coll, B., and Morel, J. M. Denoising image sequences does not require motion estimation. In *Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 70–74, Sept 2005b. doi: 10.1109/AVSS.2005.1577245.

Buades, A., Coll, B., and Morel, J.-M. Nonlocal image and movie denoising. *International Journal of Computer Vision*, 76(2):123–139, 2008. doi: 10.1007/s11263-007-0052-1.

Buades, A., Coll, B., and Morel, J.-M. Non-Local Means Denoising. *Image Processing On Line*, 1, 2011. doi: 10.5201/ipol.2011.bcm_nlm.

Champagnat, F., Le Besnerais, G., and Kulcsár, C. Statistical performance modeling for superresolution: a discrete data-continuous reconstruction framework. *Journal of the Optical Society of America A*, 26(7):1730–1746, 2009. doi: 10.1364/JOSAA.26.001730.

Chatterjee, P. and Milanfar, P. Patch-Based Near-Optimal Image Denoising. *IEEE Transactions on Image Processing*, 21(4):1635–1649, 2012. doi: dx.doi.org/10.1109/TIP.2011.2172799.

Choi, J., Park, M. K., and Kang, M. G. High Dynamic Range Image Reconstruction with Spatial Resolution Enhancement. *The Computer Journal*, 52(1):114–125, 2009. doi: 10.1093/comjnl/bxm080.

Debevec, P. E. and Malik, J. Recovering High Dynamic Range Radiance Maps from Photographs. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH, pages 369–378, 1997. doi: 10.1145/258734.258884.

Deledalle, C.-A., Denis, L., and Tupin, F. How to Compare Noisy Patches? Patch Similarity Beyond Gaussian Noise. *International Journal of Computer Vision*, 99(1):86–102, 2012. doi: 10.1007/s11263-012-0519-6.

Dempster, A. P., Laird, N. M., and Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.

Devlin, K., Chalmers, A., Wilkie, A., and Purgathofer, W. STAR Report on Tone Reproduction and Physically Based Spectral Rendering. In *Eurographics 2002*, pages

Durand, F. and Dorsey, J. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics*, 21(3):257–266, 2002. doi: 10.1145/566654.566574.

DxO Labs. `http://www.dxomark.com/Cameras/Ratings/Landscape`. Last accessed: 04/04/2014.

Efros, A. A. and Leung, T. K. Texture Synthesis by Non-Parametric Sampling. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pages 1033–, 1999. doi: dx.doi.org/10.1109/ICCV.1999.790383.

Farsiu, S., Robinson, M. D., Elad, M., and Milanfar, P. Fast and robust multiframe super resolution. *IEEE Transactions on Image Processing*, 13(10):1327–1344, 2004. doi: 10.1109/TIP.2004.834669.

Fattal, R., Lischinski, D., and Werman, M. Gradient Domain High Dynamic Range Compression. *ACM Transactions on Graphics*, 21(3):249–256, 2002. doi: 10.1145/566654.566573.

Ferradans, S., Bertalmio, M., Provenzi, E., and Caselles, V. An Analysis of Visual Adaptation and Contrast Perception for Tone Mapping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10):2002–2012, 2011. doi: 10.1109/TPAMI.2011.46.

Ferwerda, J. A., Pattanaik, S. N., Shirley, P., and Greenberg, D. P. A Model of Visual Adaptation for Realistic Image Synthesis. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH, pages 249–258, 1996. doi: 10.1145/237170.237262.

Foi, A., Trimeche, M., Katkovnik, V., and Egiazarian, K. Practical Poissonian-Gaussian Noise Modeling and Fitting for Single-Image Raw-Data. *IEEE Transactions on Image Processing*, 17:1737–1754, 2008. doi: 10.1109/TIP.2008.2001399.

Fuji Film. `https://www.fujifilmusa.com/products/digital_cameras/exr/features/page_02.html`. Last accessed: 15/04/2014.

Gallo, O., Gelfandz, N., Chen, W., Tico, M., and Pulli, K. Artifact-free high dynamic range imaging. In *Proceedings of IEEE International Conference on Computational Photography (ICCP)*, pages 1–7, 2009. doi: dx.doi.org/10.1109/ICCPHOT.2009.5559003.

Granados, M., Ajdin, B., Wand, M., Theobalt, C., Seidel, H. P., and Lensch, H. P. A. Optimal HDR reconstruction with linear digital cameras. `http://www.mpi-inf.mpg.de/~granados/projects/opthdr/index.html`. Last accessed: 15/08/2012.

Granados, M., Ajdin, B., Wand, M., Theobalt, C., Seidel, H. P., and Lensch, H. P. A. Optimal HDR reconstruction with linear digital cameras. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 215–222, 2010. doi: dx.doi.org/10.1109/CVPR.2010.5540208.

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

Grossberg, M. and Nayar, S. High Dynamic Range from Multiple Images: Which Exposures to Combine? In *Proceedings of the Workshop on Color and Photometric Methods in Computer Vision (CPMCV)*, 2003.

Gunturk, B. and Gevrekci, M. High-resolution image reconstruction from multiple differently exposed images. *IEEE Signal Processing Letters*, 13(4):197 – 200, 2006. doi: 10.1109/LSP.2005.863693.

Guo, X., Qi, X., and Harris, J. A Time-to-First-Spike CMOS Image Sensor. *IEEE Sensors Journal*, 7(8):1165–1175, 2007. doi: 10.1109/JSEN.2007.900937.

Hamilton, J. and Adams, J. Adaptive color plan interpolation in single sensor color electronic camera. US Patent 5,629,734, 1997.

Hasinoff, S., Durand, F., and Freeman, W. Noise-optimal capture for high dynamic range photography. `http://people.csail.mit.edu/hasinoff/hdrnoise/`. Last accessed: 17/03/2014.

Hasinoff, S. W., Durand, F., and Freeman, W. T. Noise-optimal capture for high dynamic range photography. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 553–560, 2010. doi: dx.doi.org/10.1109/CVPR.2010.5540167.

Healey, G. and Kondepudy, R. Radiometric CCD camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3):267 –276, 1994. doi: dx.doi.org/10.1109/34.276126.

Heo, Y., Lee, K., Lee, S., Moon, Y., and Cha, J. Ghost-free high dynamic range imaging. *Proceedings of the Asian Conference on Computer Vision (ACCV)*, pages 486–500, 2011.

Hirakawa, K. and Simon, P. Single-shot high dynamic range imaging with conventional camera hardware. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pages 1339–1346, 2011. doi: 10.1109/ICCV.2011.6126387.

Hirakawa, K. and Wolfe, P. J. Optimal exposure control for high dynamic range imaging. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 3137–3140, 2010. doi: dx.doi.org/10.1109/ICIP.2010.5654059.

Irie, K., McKinnon, A. E., Unsworth, K., and Woodhead, I. M. A model for measurement of noise in CCD digital-video cameras. *Measurement Science and Technology*, 19(4), 2008. doi: 10.1088/0957-0233/19/4/045207.

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

Jobson, D., Rahman, Z.-u., and Woodell, G. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing*, 6(7):965–976, 1997. doi: 10.1109/83.597272.

Kang, S., Uyttendaele, M., Winder, S., and Szeliski, R. High dynamic range video. *ACM Transactions on Graphics*, 22(3):319–325, 2003. doi: 10.1145/882262.882270.

Kavusi, S. and El Gamal, A. Quantitative study of high-dynamic-range image sensor architectures, 2004.

Kay, S. M. *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993. ISBN 0-13-345711-7.

Khan, E., Akyuz, A., and Reinhard, E. Ghost removal in high dynamic range images. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 2005–2008, 2006. doi: dx.doi.org/10.1109/ICIP.2006.312892.

Kirk, K. and Andersen, H. J. Noise Characterization of Weighting Schemes for Combination of Multiple Exposures. In *British Machine Vision Association (BMVC)*, pages 1129–1138, 2006.

Lebrun, M., Buades, A., and Morel, J. A nonlocal bayesian image denoising algorithm. *SIAM Journal on Imaging Sciences*, 6(3):1665–1688, 2013. doi: 10.1137/120874989.

Lyu, S. and Simoncelli, E. Modeling Multiscale Subbands of Photographic Images with Fields of Gaussian Scale Mixtures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(4):693–706, 2009. doi: 10.1109/TPAMI.2008.107.

Mancini, R. *Op Amps for everyone*. Texas Instruments, 2002.

Mann, S. and Picard, R. W. On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proceedings of IS&T*, pages 442–448, 1995.

Mantiuk, R., Daly, S., and Kerofsky, L. Display adaptive tone mapping. *ACM Transactions on Graphics*, 27(3):68:1–68:10, 2008. doi: 10.1145/1360612.1360667.

Martinec, E. Noise, dynamic range and bit depth in digital slrs. pattern noise. `http://theory.uchicago.edu/~ejm/pix/20d/tests/noise/#patternnoise`, a. Last accessed: 03/08/2012.

Martinec, E. Noise, dynamic range and bit depth in digital slrs. thermal noise. `http://theory.uchicago.edu/~ejm/pix/20d/tests/noise/#shotnoise`, b. Last accessed: 20/01/2014.

Media, X. T. Xiph.org test media. `http://media.xiph.org/sintel/sintel-4k-tiff16/`. Last accessed: 15/08/2012.

Milanfar, P. *Super-resolution imaging*, volume 1. CRC Press, 2010.

Min, T., Park, R., and Chang, S. Histogram based ghost removal in high dynamic range images. In *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, pages 530–533, 2009. doi: dx.doi.org/10.1109/ICME.2009.5202550.

Mitsunaga, T. and Nayar, S. K. Radiometric self calibration. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1374–1380, 1999. doi: http://doi.ieeecomputersociety.org/10.1109/CVPR.1999.786966.

Nakai, H., Yamamoto, S., Ueda, Y., and Shigeyama, Y. High Resolution and High Dynamic Range Image Reconstruction from Differently Exposed Images. In *Proceedings of the International Symposium on Advances in Visual Computing (ISVC)*, pages 713–722, 2008. doi: 10.1007/978-3-540-89646-3_70.

Narasimhan, S. G. and Nayar, S. K. Enhancing resolution along multiple imaging dimensions using assorted pixels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):518–530, 2005. doi: 10.1109/TPAMI.2005.76.

Nayar, S. and Mitsunaga, T. High Dynamic Range Imaging: Spatially Varying Pixel Exposures. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 472–479, Jun 2000. doi: dx.doi.org/10.1109/CVPR.2000.855857.

Papoulis, A. Generalized sampling expansion. *Circuits and Systems, IEEE Transactions on*, 24(11):652–654, 1977. doi: 10.1109/TCS.1977.1084284.

Pattanaik, S. and Yee, H. Adaptive gain control for high dynamic range image display. In *Proceedings of the Spring Conference on Computer Graphics*, SCCG, pages 83–87, 2002. doi: 10.1145/584458.584472.

Pattanaik, S. N., Ferwerda, J. A., Fairchild, M. D., and Greenberg, D. P. A Multiscale Model of Adaptation and Spatial Vision for Realistic Image Display. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH, pages 287–298, 1998. doi: 10.1145/280814.280922.

Pattanaik, S. N., Tumblin, J., Yee, H., and Greenberg, D. P. Time-dependent visual adaptation for fast realistic image display. In *Proceedings of theAnnual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH, pages 47–54, 2000. doi: 10.1145/344779.344810.

Pece, F. and Kautz, J. Bitmap movement detection: HDR for dynamic scenes. In *Proceedings of the Conference on Visual Media Production (CVMP)*, pages 1–8, 2010. doi: dx.doi.org/10.1109/CVMP.2010.8.

Pérez, P., Gangnet, M., and Blake, A. Poisson image editing. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH, pages 313–318, 2003. doi: 10.1145/1201775.882269.

Rahman, Z.-u., Jobson, D. J., and Woodell, G. A. Multiscale retinex for color rendition and dynamic range compression, 1996.

Reinhard, E. and Devlin, K. Dynamic range reduction inspired by photoreceptor physiology. *IEEE Transactions on Visualization and Computer Graphics*, 11(1):13–24, 2005. doi: 10.1109/TVCG.2005.9.

Reinhard, E., Ward, G., Pattanaik, S. N., and Debevec, P. E. *High Dynamic Range Imaging - Acquisition, Display, and Image-Based Lighting*. Morgan Kaufmann, 2005.

Reinhard, E., Ward, G., Pattanaik, S. N., Debevec, P. E., Heidrich, W., and Myszkowski, K. *High Dynamic Range Imaging - Acquisition, Display, and Image-Based Lighting*. Morgan Kaufmann, 2010.

Reinhard, E., Stark, M., Shirley, P., and Ferwerda, J. Photographic Tone Reproduction for Digital Images. *ACM Transactions on Graphics*, 21(3):267–276, 2002. doi: 10.1145/566654.566575.

Rice, J. A. *Mathematical Statistics and Data Analysis*. Duxbury Press, 1995.

Robertson, M. A., Borman, S., and Stevenson, R. L. Estimation-theoretic approach to dynamic range enhancement using multiple exposures. *Journal of Electronic Imaging*, 12(2):219–228, 2003. doi: dx.doi.org/10.1117/1.1557695.

Robinson, D. and Milanfar, P. Statistical performance analysis of super-resolution. *Image Processing, IEEE Transactions on*, 15(6):1413–1428, June 2006. ISSN 1057-7149. doi: 10.1109/TIP.2006.871079.

Schlick, C. Quantization Techniques for Visualization of High Dynamic Range Pictures. In *Photorealistic Rendering Techniques*, Focus on Computer Graphics, pages 7–20. Springer Berlin Heidelberg, 1995. doi: 10.1007/978-3-642-87825-1_2.

Schöberl, M., Belz, A., Seiler, J., Foessel, S., and Kaup, A. High dynamic range video by spatially non-regular optical filtering. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pages 2757–2760, 2012a. doi: 10.1109/ICIP.2012.6467470.

Schöberl, M., Belz, A., Nowak, A., Seiler, J., Kaup, A., and Foessel, S. Building a high dynamic range video sensor with spatially nonregular optical filtering. volume 8499, pages 84990C–84990C–11, 2012b. doi: 10.1117/12.928858.

Seiler, J. and Kaup, A. Complex-Valued Frequency Selective Extrapolation for Fast Image and Video Signal Extrapolation. *IEEE Signal Processing Letters*, 17(11):949–952, 2010. doi: 10.1109/LSP.2010.2078504.

Sen, P., Kalantari, N. K., Yaesoubi, M., Darabi, S., Goldman, D. B., and Shechtman, E. Robust patch-based HDR reconstruction of dynamic scenes. *ACM Transactions on Graphics*, 31(6):203:1–203:11, 2012. ISSN 0730-0301. doi: 10.1145/2366145.2366222.

Sidibé, D., Puech, W., and Strauss, O. Ghost detection and removal in high dynamic range images. In *Proceedings of the European Signal Processing Conference*, 2009. doi: dx.doi.org/10.1016/j.image.2012.02.001.

Srikantha, A. and Sidibé, D. Ghost detection and removal for high dynamic range images: Recent advances. *Signal Processing: Image Communication*, 2012.

Theuwissen, A. J. P. Solid-State Imaging with Charge-Coupled Devices. pages 94–108, 1996.

Tian, J. and Ma, K.-K. A survey on super-resolution imaging. *Signal, Image and Video Processing*, 5(3):329–342, 2011. ISSN 1863-1703. doi: 10.1007/s11760-010-0204-6.

Traonmilin, Y. and Aguerrebere, C. Simultaneous high dynamic range and super-resolution imaging without regularization. *SIAM Journal on Imaging Science (to appear)*, 2014.

Traonmilin, Y., Ladjal, S., and Almansa, A. On the Amount of Regularization for Super-Resolution Interpolation. In *Proceedings of the European Signal Processing Conference (EUSIPCO)*, 2012a.

Traonmilin, Y., Ladjal, S., and Almansa, A. On the amount of regularization for super-resolution reconstruction. *Preprint HAL* `http://hal.archives-ouvertes.fr/hal-00763984`, 2012b.

Traonmilin, Y., Ladjal, S., and Almansa, A. Outlier Removal Power of the L1-Norm Super-Resolution. In *Scale Space and Variational Methods in Computer Vision*, volume 7893 of *Lecture Notes in Computer Science*, pages 198–209. Springer Berlin Heidelberg, 2013. doi: 10.1007/978-3-642-38267-3_17.

Tsin, Y., Ramesh, V., and Kanade, T. Statistical calibration of the CCD imaging process. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pages 480–487, 2001. doi: dx.doi.org/10.1109/ICCV.2001.937555.

Tumblin, J. and Rushmeier, H. Tone Reproduction for Realistic Images. *IEEE Computer Graphics and Applications*, 13(6):42–48, 1993. doi: 10.1109/38.252554.

Tumblin, J. and Turk, G. LCIS: A Boundary Hierarchy for Detail-preserving Contrast Reduction. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH, pages 83–90, 1999. doi: 10.1145/311535.311544.

Tumblin, J., Hodgins, J. K., and Guenter, B. K. Two methods for display of high contrast images. *ACM Transactions on Graphics*, 18(1):56–94, 1999. ISSN 0730-0301. doi: 10.1145/300776.300783.

Wang, Y.-Q. E-PLE: an Algorithm for Image Inpainting. *Image Processing On Line*, 2013: 271–285, 2013a. doi: 10.5201/ipol.2013.54.

Wang, Y.-Q. The Implementation of SURE Guided Piecewise Linear Image Denoising. *Image Processing On Line*, 3:43–67, 2013b. doi: 10.5201/ipol.2013.52.

Wang, Y.-Q. and Morel, J.-M. SURE Guided Gaussian Mixture Image Denoising. *SIAM Journal on Imaging Sciences*, 6(2):999–1034, 2013. doi: dx.doi.org/10.1137/120901131.

Ward, G. A contrast-based scalefactor for luminance display. *Graphics Gems IV*, pages 415–421, 1994.

Ward, G., Rushmeier, H., and Piatko, C. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Transactions on Visualization and Computer Graphics*, 3(4):291–306, 1997. doi: 10.1109/2945.646233.

Wexler, Y., Shechtman, E., and Irani, M. Space-time completion of video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(3):463–476, 2007. doi: doi.ieeecomputersociety.org/10.1109/TPAMI.2007.60.

Yasuma, F., Mitsunaga, T., Iso, D., and Nayar, S. Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum. *IEEE Transactions on Image Processing*, 99, Mar 2010. doi: 10.1109/TIP.2010.2046811.

Yu, G., Sapiro, G., and Mallat, S. Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity. *IEEE Transactions on Image Processing*, 21(5):2481–2499, 2012. doi: 10.1109/TIP.2011.2176743.

Zimmer, H., Bruhn, A., and Weickert, J. Freehand HDR imaging of moving scenes with simultaneous resolution enhancement. *Computer Graphics Forum (Eurographics)*, 30(2):405–414, 2011. doi: dx.doi.org/10.1111/j.1467-8659.2011.01870.x.

Zoran, D. and Weiss, Y. From learning models of natural image patches to whole image restoration. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pages 479–486, 2011. doi: dx.doi.org/10.1109/ICCV.2011.6126278.