High Dynamic Range Imaging: Problems of Video Exposure Bracketing, Luminance Calibration and Gloss Editing



A dissertation submitted towards the degree
Doctor Engineering (Dr.-Ing)
of the Faculty of Mathematics and Computer Science
of Saarland University

By

Yulia Gryaditskaya

December 2016, Saarbrücken, Germany

Advisors:
Dr.-Ing. habil. Karol Myszkowski
Prof. Dr. Hans-Peter Seidel

Examination

Dean of the Faculty:
Univ.-Prof. Dr. Frank-Olaf Schreyer

Date:
02 - 06 - 2017

Chair of the Committee:
Prof. Dr. Matthias Hein

Examiners:
Dr. Adrien Bousseau
Dr.-Ing. habil. Karol Myszkowski
Dr. Erik Reinhard
Prof. Dr. Hans-Peter Seidel

Academic Assistant:
Dr. Shida Beigpour

Abstract

Two-dimensional, conventional images are gradually losing their hegemony, leaving room for novel formats. Among these, 8 bit images give place to high dynamic range (HDR) image formats, allowing to improve colour gamut and visibility of details in dark and bright areas leading to a more immersive viewing experience. It opens wide opportunities for post-processing, which can be useful for artistic rendering, enhancement of viewing experience or medical applications. Simultaneously, light-field scene representation as well is gaining importance, propelled by the recent reappearance of virtual reality, the improvement of both acquisition techniques, and computational and storage capabilities. Light-field data as well allows to achieve a broad range of effects in post-production: among others, it enables a change of a camera position, an aperture or a focal length. It facilitates object insertions and simplifies visual effects workflow by integrating 3D nature of visual effects with 3D nature of light fields.

Content generation is one of the stumbling blocks in these realms. Sensor limitations of a conventional camera do not allow to capture wide dynamic range. This especially is the case for mobile devices, where small sensors are optimised for capturing in high-resolution. The "HDR mode" often encountered on such devices, relies on techniques called "exposure fusion" and allows to partially overcome the limited range of a sensor. The HDR video at the same time remains a challenging problem. We suggest a solution for an HDR video capturing on a mobile device. We analyse dynamic range of motion regions, the regions which are the most prone to reconstruction artefacts, and suggest a real-time exposure selection algorithm. Further, an HDR content visualization task often requires an input to be in absolute values. We address this problem by presenting a calibration algorithm that can be applied to existent imagery and does not require any additional measurement hardware. Finally, as light fields use becomes more common, a key challenge is the ability to edit or modify the appearance of the objects in the light field. To this end, we propose a multidimensional filtering approach in which the specular highlights are filtered in the spatial and angular domains to target a desired increase of the material roughness.

Kurzzusammenfassung

Klassische zweidimensionale Bilder verlieren allmählich ihre Vorherrschaft und ermöglichen damit das Aufkommen neuartiger Formate. Unter anderem werden 8-Bit-Bilder von Bildern mit hohem Dynamikumfang (HDR) verdrängt, die eine Verbesserung des Farbraums und die Sichtbarkeit von Details in dunklen und hellen Bereichen möglich machen und somit zu einem immersiveren Seherlebnis führen. Dies eröffnet weitreichende Möglichkeiten für die Postproduktion, die für künstlerisches Rendering, die Verbesserung des Seherlebnisses oder medizinische Anwendungen nützlich sind. Angetrieben vom jüngsten Wiederaufleben von virtueller Realität und der Verbesserung von sowohl Aufnahmetechnologie als auch der Rechen- und Speicherleistung, gewinnt auch die Lichtfeld-Szenenrepräsentation an Bedeutung. Lichtfelddaten lassen umfangreiche Effekte in der Postproduktion zu: Unter anderem sind Änderungen von Position, Blende oder Brennweite der Kamera möglich. Objekte können nachträglich in die Szene eingefügt werden und der Workflow für visuelle Effekte wird vereinfacht, da sich dreidimensionale visuelle Effekte natürlicherweise in dreidimensionale Lichtfelder integrieren lassen.

In diesem Gebiet ist das Erzeugen von Inhalten ist einer der Stolpersteine. Sensorbeschränkungen von herkömmlichen Kameras führen dazu, dass hohe Dynamikumfänge nicht aufgezeichnet werden können. Dies ist besonders in mobilen Geräten der Fall, bei denen kleine Sensoren für hochauflösende Aufnahmen optimiert sind. Der “HDR Modus”, der in diesen Geräten häufig zu finden ist, beruht auf der sogenannten “Belichtungsfusion” und kann den eingeschränkten Umfang des Sensors teilweise erweitern. HDR-Videos stellen hingegen weiterhin ein anspruchsvolles Problem dar. Wir schlagen eine Lösung vor, um HDR-Videos auf mobilen Geräten aufzuzeichnen. Wir analysieren den Dynamikumfang von Bewegungsbereichen, also solchen Bildbereichen, die am anfälligsten für Rekonstruktionsartefakte sind, und schlagen einen Echtzeitalgorithmus zur Belichtungsauswahl vor. Darüber hinaus werden für die Visualisierung von HDR-Inhalten häufig Eingaben auf einer absoluten Skala. Wir widmen uns diesem Problem, indem wir einen Kalibrierungsalgorithmus vorschlagen, der auf bestehendes Bildmaterial angewendet werden kann und keine zusätzliche Vermessungshardware benötigt. Da Lichtfeldanwendungen immer verbreiteter werden, stellt das Bearbeiten oder Verändern des Erscheinungsbildes von Objekten in Lichtfeldern eine zentrale Herausforderung dar. Zu diesem Zweck schlagen wir einen multidimensionalen Filteransatz vor, bei dem Glanzlichter im räumlichen und im Winkelraum gefiltert werden, um die Unebenheit von Materialien kontrolliert zu verstärken.

Summary

Two-dimensional, conventional images are gradually losing their hegemony, leaving room for novel formats. Among these, 8 bit images give place to high dynamic range (HDR) image formats, allowing to improve colour gamut and visibility of details in dark and bright areas leading to a more immersive viewing experience. It opens wide opportunities for post-processing, which can be useful for artistic rendering, enhancement of viewing experience or medical applications. Simultaneously, light-field scene representation as well is gaining importance, propelled by the recent reappearance of virtual reality, the improvement of both acquisition techniques, and computational and storage capabilities. Light-field data as well allows to achieve a broad range of effects in post-production: among others, it enables a change of a camera position, an aperture or a focal length. It facilitates object insertions and simplifies visual effects workflow by integrating 3D nature of visual effects with 3D nature of light fields.

Content generation is one of the stumbling blocks in these realms. Sensor limitations of a conventional camera do not allow to capture wide dynamic range. This especially is the case for mobile devices, where small sensors are optimised for capturing in high-resolution. The "HDR mode" often encountered on such devices, relies on techniques called "exposure fusion" and allows to partially overcome the limited range of a sensor. The HDR video at the same time remains a challenging problem. We suggest a solution for an HDR video capturing on a mobile device. We analyse dynamic range of motion regions, the regions which are the most prone to reconstruction artefacts, and suggest a real-time exposure selection algorithm. Further, an HDR content visualization task often requires an input to be in absolute values: whether the target media is an HDR or a standard/low dynamic range display. In case of the latter the mapping of the HDR content to LDR content is often driven by colour appearance models, which require the input to be in absolute values. We address this problem by presenting a calibration algorithm that can be applied to existent imagery and does not require any additional measurement hardware. Finally, as light fields use becomes more common, a key challenge is the ability to edit or modify the appearance of the objects in the light field. To this end, we propose a multidimensional filtering approach in which the specular highlights are filtered in the spatial and angular domains to target a desired increase of the material roughness. To achieve this we address the problem of depth estimation in the presence of bright specular regions in the scene.

The text is structured as follows. First, the motivation is introduced in Chapter 1. In Chapters 2 and 3 we describe the background on which the thesis builds. The core contributions of the thesis are then detailed in Chapters 4–6.

Capturing Chapter 4 presents an algorithm that allows HDR video capturing with a conventional camera, such as the one that can be found on a mobile device (based on [Gryaditskaya et al. 2015, Pouli et al. 2016]). The task is important, since mobile phones and tablets are rapidly gaining significance as omnipresent image and video capture devices. The design of the algorithm was

informed by a perceptual study that assesses the relative importance of motion and dynamic range. We found that ghosting artefacts are more visually disturbing than a reduction in dynamic range, even if a comparable number of pixels is affected by each. We incorporated these findings into a real-time, adaptive metering algorithm that allows to significantly reduce artefacts in an HDR sequence obtained as a recombination of differently exposed frames. Finally, we present an off-line HDR reconstruction algorithm that can handle non-equally spaced exposure adapted to the scene content.

Calibration Chapter 5 introduces a novel solution for HDR images calibration which does not require any additional measuring equipment and can be applied to existing imagery (based on [Gryaditskaya et al. 2014]). Image calibration requires both linearisation of pixel values and scaling so that values in the image correspond to real-world luminances. In this chapter we focus on the latter and rather than rely on camera characterization, we calibrate images by analysing their content and metadata, obviating the need for expensive measuring devices or modelling of lens and camera combinations. Our analysis correlates sky pixel values to luminances that would be expected based on geographical metadata. Combined with HDR imaging, which gives us linear pixel data, our algorithm allows us to find absolute luminance values for each pixel—effectively turning digital cameras into absolute light meters. To validate our algorithm we have collected and annotated a calibrated set of HDR images and compared our estimation with several other approaches, showing that our solution is able to more accurately recover absolute luminance. We discuss various applications and demonstrate the utility of our method in the context of calibrated colour appearance reproduction and lighting design.

Editing Chapter 6 presents a method for manipulating the appearance of gloss in light fields (based on [Gryaditskaya et al. 2016]). We propose a multidimensional filtering approach in which the specular highlights are filtered in the spatial and angular domains to target a desired increase of the material roughness. The filtering kernel is computed based on surface normals and view direction. Our technique generates angularly-coherent plausible edits. We present a full pipeline, which includes depth estimation, normal vectors computation, specular layers separation and filtering. We develop a method of depth estimation in the presence of specular regions, which relies on the assumption that there exists a view where the point is observed as diffuse. We further show how to obtain angularly consistent depth estimates.

Finally, the summary and conclusions of the work are discussed in Chapter 7, including an overview of possible future work.

Zusammenfassung

Klassische zweidimensionale Bilder verlieren allmählich ihre Vorherrschaft und ermöglichen damit das Aufkommen neuartiger Formate. Unter anderem werden 8-Bit-Bilder von Bildern mit hohem Dynamikumfang (HDR) verdrängt, die eine Verbesserung des Farbraums und die Sichtbarkeit von Details in dunklen und hellen Bereichen möglich machen und somit zu einem immersiveren Seherlebnis führen. Dies eröffnet weitreichende Möglichkeiten für die Postproduktion, die für künstlerisches Rendering, die Verbesserung des Seherlebnisses oder medizinische Anwendungen nützlich sind. Angetrieben vom jüngsten Wiederaufleben von virtueller Realität und der Verbesserung von sowohl Aufnahmetechnologie als auch der Rechen- und Speicherleistung, gewinnt auch die Lichtfeld-Szenenrepräsentation an Bedeutung. Lichtfelddaten lassen umfangreiche Effekte in der Postproduktion zu: Unter anderem sind Änderungen von Position, Blende oder Brennweite der Kamera möglich. Objekte können nachträglich in die Szene eingefügt werden und der Workflow für visuelle Effekte wird vereinfacht, da sich dreidimensionale visuelle Effekte natürlicherweise in dreidimensionale Lichtfelder integrieren lassen.

In diesem Gebiet ist das Erzeugen von Inhalten ist einer der Stolpersteine. Sensorbeschränkungen von herkömmlichen Kameras führen dazu, dass hohe Dynamikumfänge nicht aufgezeichnet werden können. Dies ist besonders in mobilen Geräten der Fall, bei denen kleine Sensoren für hochauflösende Aufnahmen optimiert sind. Der “HDR Modus”, der in diesen Geräten häufig zu finden ist, beruht auf der sogenannten “Belichtungsfusion” und kann den eingeschränkten Umfang des Sensors teilweise erweitern. HDR-Videos stellen hingegen weiterhin ein anspruchsvolles Problem dar. Wir schlagen eine Lösung vor, um HDR-Videos auf mobilen Geräten aufzuzeichnen. Wir analysieren den Dynamikumfang von Bewegungsbereichen, also solchen Bildbereichen, die am anfälligsten für Rekonstruktionsartefakte sind, und schlagen einen Echtzeitalgorithmus zur Belichtungsauswahl vor. Darüber hinaus werden für die Visualisierung von HDR-Inhalten häufig Eingaben auf einer absoluten Skala benötigt, unabhängig davon, ob das Zielmedium ein HDR- oder ein Standarddisplay mit niedrigem Dynamikumfang ist. In letzterem Fall wird die Abbildung von HDR- auf LDR-Inhalte oft von Farbdarstellungsmodellen gesteuert, die absolute Skalen als Eingabe benötigen. Wir widmen uns diesem Problem, indem wir einen Kalibrierungsalgorithmus vorschlagen, der auf bestehendes Bildmaterial angewendet werden kann und keine zusätzliche Vermessungshardware benötigt. Da Lichtfeldanwendungen immer verbreiteter werden, stellt das Bearbeiten oder Verändern des Erscheinungsbildes von Objekten in Lichtfeldern eine zentrale Herausforderung dar. Zu diesem Zweck schlagen wir einen multidimensionalen Filteransatz vor, bei dem Glanzlichter im räumlichen und im Winkelraum gefiltert werden, um die Unebenheit von Materialien kontrolliert zu verstärken. Um dies zu erreichen, behandeln wir das Problem der Tiefenbestimmung bei hellen spiegelnden Bereichen in der Szene.

Der Text ist wie folgt gegliedert. Zuerst werden in Kapitel 1 die Motivation vorgestellt. In den Kapiteln 2 and 3 werden die Grundlagen beschrieben, auf denen diese Arbeit aufgebaut ist. Daraufhin werden in den Kapiteln 4–6 die Kernbeiträge dieser Arbeit ausgeführt.

Aufzeichnung Kapitel 4 stellt einen Algorithmus vor, der es ermöglicht, HDR-Videos mit einer gewöhnlichen, beispielsweise in ein mobiles Gerät eingebauten Kamera aufzuzeichnen (basierend auf [Gryaditskaya et al. 2015, Pouli et al. 2016]). Dies stellt eine wichtige Anwendung dar, da Handys und Tablets als allgegenwärtige Geräte zur Foto- und Videoaufzeichnung einen immer größeren Stellenwert haben. Dem Design des Algorithmus' liegt eine Wahrnehmungsstudie zu Grunde, mit deren Hilfe die relative Gewichtung von Bewegung und Dynamikumfang beurteilt werden kann. Wir können feststellen, dass Ghosting-Artefakte den Seheindruck stärker negativ beeinflussen als ein reduzierter Dynamikumfang, selbst wenn die Anzahl der jeweils betroffenen Pixel vergleichbar ist. Diese Erkenntnisse werden in einem adaptiven Echtzeit-Vermessungsalgorithmus verwendet, der Artefakte in HDR-Bildsequenzen, die durch Rekombination von verschieden belichteten Einzelbildern erzeugt werden, erheblich reduziert. Schließlich stellen wir einen offline arbeitenden HDR-Rekonstruktionsalgorithmus vor, der mit ungleichmäßig platzierten Belichtungen, die sich an den Szeneninhalt anpassen, umgehen kann.

Kalibrierung Kapitel 5 stellt eine neuartige Lösung zum Kalibieren von HDR-Bildern vor, die ohne zusätzliche Vermessungshardware auskommt und auf bestehendes Bildmaterial angewendet werden kann (basierend auf [Gryaditskaya et al. 2014]). Damit die abgebildeten Größen echten Leuchtdichtewerten entsprechen, erfordert die Bildkalibrierung sowohl eine Linearisierung als auch eine Skalierung der Pixelwerte. In diesem Kapitel betrachten wir Letzteres und kalibrieren Bilder mit Hilfe einer Analyse von Bildinhalt und Metadaten, statt eine Kameracharakterisierung durchzuführen. Auf diese Weise vermeiden wir kostspielige Vermessungsgeräte und das Modellieren von Kombinationen aus Linse und Kamera. Unsere Analyse korreliert Pixelwerte des Himmels mit Leuchtdichten, die auf Grund von geografischen Metadaten zu erwarten sind. In Kombination mit HDR-Bildgebungsmethoden, die lineare Pixeldaten liefern, errechnet unser Algorithmus absolute Leuchtdichtewerte für jeden Pixel und verwandelt digitale Kameras damit in absolute Belichtungsmesser. Um unseren Algorithmus zu validieren wird eine Sammlung kalibrierter HDR-Bilder zusammengetragen und annotiert, und unsere Abschätzung mit mehreren anderen Ansätzen verglichen. Es stellt sich heraus, dass unsere Lösung in der Lage ist, genauere absolute Leuchtdichtewerte zu ermitteln. Wir diskutieren diverse Anwendungen und zeigen den Nutzwert unserer Methode im Zusammenhang mit der Wiedergabe kalibrierter Farben und beim Beleuchtungsdesign.

Bearbeitung Kapitel 6 beschreibt eine Methode, um das Erscheinungsbild von Glanzlichtern in Lichtfeldern zu manipulieren (basierend auf [Gryaditskaya et al. 2016]). Wir schlagen einen multidimensionalen Filteransatz vor, bei dem Glanzlichter im räumlichen und im Winkelraum gefiltert werden, um die Unebenheit von Materialien kontrolliert zu verstärken. Der Filterkernel wird dabei auf der Grundlage von Oberflächennormalen und der Blickrichtung berechnet. Unser Verfahren erzeugt plausible Resultate und ist im Winkelraum kohärent. Wir stellen eine komplette Pipeline mit Tiefenbestimmung, Berechnung von Normalvektoren, Trennung von Glanzebenen und Filterung vor. Wir entwickeln eine Methode zur Tiefenbestimmung bei spiegelnden Bereichen, die auf der Annahme beruht, dass eine Ansicht existiert, unter welcher der darzustellende Punkt diffus erscheint. Weiterhin zeigen wir, wie winkelkonsistente Tiefenbestimmungen erreicht werden können.

Zum Schluss wird in Kapitel 7 eine Zusammenfassung gegeben und ein Fazit gezogen. Dabei wird ein überblick über möglichen weiteren Forschungsbedarf vermittelt.

Acknowledgments

First of all, I would like to thank Hans-Peter Seidel and Karol Myszkowski for giving me an opportunity to become a member of the computer graphics research group at Max Planck Institute. It was a pleasure to work among very talented and motivated people, to become friends with many of them. The work conditions as well as people attitude to work at MPI set for me a high standard.

Over the time spent there I learnt so much: not only in terms of obtained knowledge and skills, I also learned a lot about myself. This would be impossible without so many people whom I would like to acknowledge here.

I am sincerely grateful to Erik Reinhard and Tania Pouli for taking time and patience to open a research world for me. Your approach to solving problems and finding solutions, your energy encourage me. Moreover, I truly thankful for all the great time we spent outside work.

I eternally grateful to Karol Myszkowski for all the opportunities: to work with you and other inspiring people, to work on the things which were truly interesting for me; For giving me time to learn and explore, and to develop myself; For always warm working atmosphere and trust.

I further was truly lucky to work with Piotr Didyk and Belen Masia. Thank you guys for a very enjoyable work process; For finding time for me, independently of how busy your calendars are; For multiple iterations over my slides and the paper, for explaining me why and what. And again, both of you, thank you for being to me more than colleagues.

Thanks to all the members of the group, it was really great to be colleagues with you. I especially thank Petr Kellnhofer for being a friend whom I can rely on, for all the time we spent together.

I want to thank Oskar Elek, for even not letting me to assume that I can not do something; for pushing me out of my comfort zone; for telling me your opinion faithfully, providing the critics and acknowledgement equally; for all the discussions and advices. And of course, thank you for bringing to my life running and biking, for pull ups at 2am, it was a lot of fun.

Thank you all my dear friends, with a special acknowledgement of Monica Vidriales, Himangshu Saikia, Gregorio Palmas, Lisette Noboa, Mateusz Malinowski, Krzysztof Templin, Daria Gaidar and Daniyar Kurmankhojayev, for making rainy days in Sarrbruecken truly pleasant and unforgettable.

I thank all my friends who are staying with me independently of how far geographically I'm; your believe in me, support and love keep me going. Among others I want to distinguish: Anna Laskovaya, Irina Spirina and Larisa Chodorenko (Shevchenko). I would like to separately thank Gregory Bonik who influenced what I'm and taught me a lot.

Finally, I thank my parents who always believe in me, support everything I choose to do, independently on how steep I turn; Who let me make my own choices and decisions from early childhood; Who gave me the best possible education. And the most important thank you for your endless love.

Contents

1 Introduction **1**
- 1.1 Motivation . . . 1
- 1.2 Contributions and Structure . . . 2

2 Background **5**
- 2.1 HDR Imaging . . . 5
- 2.2 Dynamic Range Limitations in Conventional Capturing Systems . . . 6
- 2.3 HDR Imaging Acquisition . . . 7
  - 2.3.1 Specialised Cameras . . . 8
  - 2.3.2 Multi-Exposure Techniques for Images . . . 9
  - 2.3.3 Multi-Exposure Techniques for Videos . . . 10
  - 2.3.4 Summary . . . 11
- 2.4 Calibration . . . 11
  - 2.4.1 Scene Measurements . . . 11
  - 2.4.2 Camera Characterization . . . 11
  - 2.4.3 EXIF Based Approach . . . 12
  - 2.4.4 Summary . . . 16
- 2.5 Light Fields . . . 17
- 2.6 Registration Problem in Light Fields . . . 17
  - 2.6.1 Epipolar Plane Images Analysis . . . 18
  - 2.6.2 Full 4D Angular Consistency . . . 20
  - 2.6.3 Depth Estimation for Non-Lambertian Surfaces . . . 22
  - 2.6.4 Summary . . . 24
- 2.7 HDR Light Field . . . 24

3 Related Work **27**
- 3.1 Perception of Ghosting Artefacts . . . 27
- 3.2 Perception of Gloss . . . 28
  - 3.2.1 Gloss Consistency Between Different Views . . . 28
- 3.3 Material Editing . . . 28

4 Motion Aware Exposure Bracketing For HDR Video **31**
- 4.1 Perceptual Assessment of Ghosting and Over-Exposure Artefacts . . . 32
  - 4.1.1 Stimuli . . . 33
  - 4.1.2 Experimental Setup . . . 33
  - 4.1.3 Analysis . . . 34
- 4.2 Metering Algorithm . . . 35
  - 4.2.1 Initialization . . . 36
  - 4.2.2 Motion Estimation . . . 36

4.3 Histogram Computation . . . 37
4.3.1 Exposure Time Selection . . . 38
4.3.2 Exposure Time Direction . . . 41
4.3.3 Thresholds . . . 42
4.3.4 Discussion . . . 42
4.3.5 Implementation . . . 43
4.4 HDR Video Reconstruction . . . 43
4.5 Results and Limitations . . . 45

5 Sky Based Light Metering for High Dynamic Range Images 51
5.1 Algorithm . . . 53
5.1.1 Sky Appearance . . . 54
5.1.2 The Fitting to Parametric Sky Model . . . 56
5.1.3 Sky Segmentation and Classification . . . 57
5.1.4 Relative Zenith Luminance . . . 58
5.1.5 Absolute Zenith Luminance . . . 59
5.2 Scale Factor Estimation for Angular Maps . . . 60
5.3 Ground Truth Dataset . . . 61
5.4 Algorithm Validation . . . 61
5.4.1 Clear-Sky Images . . . 61
5.4.2 Overcast Sky Images . . . 65
5.4.3 Validation of Scale Factors for Angular Maps . . . 66
5.5 The Hosek and Wilkie Sky Model . . . 66
5.6 Discussion and Limitations . . . 68

6 Gloss Editing in Light Fields 69
6.1 Light Field Preprocessing . . . 70
6.1.1 Depth Estimation . . . 70
6.1.2 Normals Computation . . . 74
6.1.3 Specular and Diffuse Component Separation . . . 74
6.2 Filtering . . . 74
6.2.1 Filtering Parameters and Validation . . . 76
6.3 Results and Limitations . . . 79

7 Conclusion 83
7.1 Summary and Outlooks . . . 83
7.1.1 HDR Video Capturing . . . 84
7.1.2 Calibration . . . 84
7.1.3 Gloss Editing . . . 85
7.2 Future Work . . . 86

Chapter 1

Introduction

This thesis focuses on capturing, calibration and editing of new scene representation formats that capture much richer information about a scene, compared to standard single view 8-bit images or videos. The formats considered in this work are high dynamic range (HDR) images and videos as well as light fields. The presented algorithms are largely motivated by the appearance of new capturing and displaying devices that allow much more faithful reproduction of a scene than the conventional media. The new displays boost the development of the content acquisition and generation techniques. Apart from the development of displaying and capturing technologies our work is motivated by a fast increasing interest in virtual reality.

In the first chapter, we explain the motivation for our work, and outline the thesis structure and its contributions.

1.1 Motivation

The history of photography is starting from discovering the camera obscura phenomenon and the property of some chemical substances to be altered by being exposed to light. The first attempts of capturing a scene using automatic processes date back to the beginning of the 19th century. Since its invention the photography pioneers were suggesting different solutions to improve the capturing capabilities. For instance, multiple solutions for colour photography were proposed as experimental setups until these solutions started to become commercially available at the end of the 19th century. Moreover, already in the mid 19th century the first attempts to expand the dynamic range using the combination of two differently exposed negatives were made. Almost one hundred years after the first success in automatic capture of a scene, the digital photography was invented and nowadays it almost fully substituted the film photography. Since then, the imaging technologies and video technologies were actively developing. Nevertheless, we are still far from being able to recreate the real scene experience not only in terms of the real-world dimensionality, spatial and temporal resolutions, but as well in terms of the dynamic range and the colour gamut which can be captured and reproduced.

Since the invention of an 8-bit CCD in 1970, this bit depth remains a standard for the popular image formats (e.g. JPEG and PNG) due to limitations on the dynamic range that can be printed and displayed. While the sensors of DSLR cameras are capable of capturing the scene content with 10-16 bits per colour channel, many mobile devices are still offering just 24-bit colour depth.

Just recently the display manufacturers started to offer 10-bits colour channels as well as improved contrast. On the other hand, improvement of technologies allows to produce true High Dynamic Range (HDR) as well as light-field displays, nevertheless still limited to mass-production by its high-cost.

There are a number of reasons to believe that HDR and light field are the key technologies of the future. One of the main reasons is that the development of both technologies is propelled by its large potential in cinematic and entertainment industries. HDR imaging provides extended colour gamut and improved contrast. It allows to render perceptual effects such as loss of colour vision in dim conditions or increased colourfulness in brightly illuminated scenes. HDR colour grading is one of key developing branches in post-production companies. There, the focus is additionally lying on generating HDR content from conventional Low Dynamic Range (LDR) content. Further, HDR content allows to render LDR content matching the dynamic range of the conventional displays, in a way, that when observed by a viewer, the perception of the image was similar to the one of the captured scene [Reinhard et al. 2010, 2012]. Light-field representation of a scene allows for each scene point to get the information not only about colour, but as well its position and light direction. Given the completeness of light fields, they have a potential of becoming the ultimate visual data representation for digital imaging. Thus, representing a scene as a light field potentially enables reproducing all visual cues that can be observed in the real world, including motion parallax and accommodation, which are usually missing in conventional imaging pipelines. It solves the problem, which the visual effects artists traditionally have to deal with: the discrepancy between three dimensional nature of visual effects and two dimensionality of the video sequences. Its increasing popularity is confirmed by the active development of capturing technologies, both in academia and industry: LYTRO recently produced the first solution for cinema, Fraunhofer as well presented a setup for light-field video capturing and demonstrated a number of effects achievable in post-production.

1.2 Contributions and Structure

In this section we briefly summarise the contributions presented in this thesis and describe its general structure. The thesis considers the problems of HDR content capturing and calibration as well as editing in light fields. Editing exploits the advantage of HDR content and at the same time solves the challenging problem of images registration in the presence of bright specular regions, characteristic for scenes with high dynamic range.

In Chapter 2 we describe the background on which the thesis builds. We first give an introduction into HDR imaging, discuss limitations on the dynamic range of conventional camera sensors as well as describe state-of-the-art in HDR content capturing technologies. We additionally discuss how an HDR content can be calibrated to store the actual scene radiance values. We then introduce the representation of a scene with a light field and give an overview of depth estimation approaches which make use of information implicitly encoded in a light field. We conclude the chapter with a survey of the HDR light-field-capturing setups.

We give an overview of related work on perception and material editing in Chapter 3.

The core part, Chapters 4–6, describes several novel methods, addressing our main contribution. This part is based on the following publications: [Gryaditskaya et al. 2014, 2015, Pouli et al. 2016, Gryaditskaya et al. 2016].



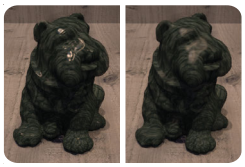
Chapter 4 presents an algorithm that allows HDR video capturing with a conventional camera, such as the one that can be found on a mobile device (based on [Gryaditskaya et al. 2015, Pouli et al. 2016]). The task is important, since mobile phones and tablets are rapidly gaining significance as omnipresent image and video capture devices. The design of the algorithm was informed by a perceptual study that assesses the relative importance of motion and dynamic range. We found that ghosting artefacts are more visually disturbing than a reduction in dynamic range, even if a comparable number of pixels is affected by each. We incorporated these findings into a real-time, adaptive metering algorithm that allows to significantly reduce artefacts in an HDR sequence obtained as a recombination of differently exposed frames. Finally, we present an off-line HDR reconstruction algorithm that can handle non-equally spaced, adapted to the scene content, exposures. In summary, we present the following contributions:

- We show for the first time a system that allows efficient metering in the context of HDR video capture, taking into account both luminance and motion.
- The algorithm is tuned on the basis of a fundamental psychophysical experiment.
- We describe an HDR video reconstruction algorithm that takes an arbitrarily varying spacing of LDR exposures and produces high quality HDR video without losing frame-rate.



Chapter 5 introduces a novel solution for HDR images calibration which does not require any additional measuring equipment and can be applied to existing imagery (based on [Gryaditskaya et al. 2014]). Image calibration requires both linearisation of pixel values and scaling so that values in the image correspond to real-world luminances. In this chapter we focus on the latter and rather than rely on camera characterization, we calibrate images by analysing their content and metadata, obviating the need for expensive measuring devices or modelling of lens and camera combinations. Our analysis correlates sky pixel values to luminances that would be expected based on geographical metadata. Combined with HDR imaging, which gives us linear pixel data, our algorithm allows us to find absolute luminance values for each pixel—effectively turning digital cameras into absolute light meters. To validate our algorithm we have collected and annotated a calibrated set of HDR images and compared our estimation with several other approaches, showing that our solution is able to more accurately recover absolute luminance. We discuss various applications and demonstrate the utility of our method in the context of calibrated colour appearance reproduction and lighting design. In summary, this work offers the following contributions and benefits:

- No special equipment is required, nor is there a need to characterize the camera or lens.
- No EXIF or GPS data is required for environment maps with a fully visible sky dome.
- Our algorithm enables the recovery of absolute values with accuracy comparable to what could be obtained from direct scene measurements.
- A calibrated and annotated HDR dataset of outdoor scenes, which includes GPS and EXIF metadata [Gryaditskaya et al. 2014].



Chapter 6 presents a method for manipulating the appearance of gloss in light fields (based on [Gryaditskaya et al. 2016]). We propose a multidimensional filtering approach in which the specular highlights are filtered in the spatial and angular domains to target a desired increase of the material roughness. The filtering kernel is computed based on surface normals and view direction. Our technique generates angularly-coherent plausible edits. We describe a full pipeline, which includes depth estimation, normal vectors computation, specular layers separation and filtering. We develop a method of depth estimation in the presence of specular regions, which relies on the assumption that there exists a view where the point is observed as diffuse. We further show how to obtain angularly consistent depth estimates. In summary, we present the following contributions:

- We demonstrate the first algorithm for consistent gloss editing in light fields.
- We describe the full pipeline, starting from depth estimation and normal computation to specular septation and editing.
- We formulate editing as a filtering problem and explore filtering in different dimensions.

The thesis is concluded in Chapter 7 by providing a summary of the work. Further directions and future outlooks are also discussed for each of the presented topics and their combinations.

Chapter 2

Background

In this chapter we give an introduction into High Dynamic Range (HDR) imaging and video (Section 2.1). We discuss the problem of limited dynamic range of conventional capturing devices (Section 2.2). We then present an overview of specialized cameras for HDR content capture as well as algorithmic solutions facilitating HDR content capture with conventional cameras (Section 2.3). Further, in Section 2.4, we discuss different approaches which allow to match HDR content to the actual radiance values of a scene.

We as well introduce light-field scene representation (Section 2.5) and discuss the problem of depth estimation from a light-field input (Section 2.6). We terminate this chapter with an overview of existing work on HDR light field capturing in Section 2.7.

2.1 HDR Imaging

HDR imaging allows to capture a much wider colour gamut, as well as luminance and contrast ranges, that are possible to obtain with traditional imaging. In order to understand the limitations of conventional imaging let us consider an example in Figure 2.1: The typical outdoor scene contains both very bright and very dark areas which cannot be captured simultaneously. At the same time, the human visual system functions over a range of about five orders of luminance magnitude simultaneously. Moreover, it is capable of adapting to lighting conditions that vary by nearly ten orders of magnitude [Ferwerda 2001]. Apart from perceptual discrepancy, missing information poses limitations on attainable image processing.

Therefore, the images obtained with conventional technologies, such as, for example, 24-bit sRGB images, are referred to as Low Dynamic Range (LDR) images to emphasize the limited dynamic range. While the images that store a depiction of a scene in a range of intensities corresponding to physical luminance or spectral radiance values are called HDR images.

Traditional LDR image formats (such as JPEG and PNG) have been designed to comply with the capabilities of display devices and thus are called *device-referred* formats. By its definition, HDR images record a much wider colour gamut than conventional LDR images and, thus, require an efficient encoding to avoid an excess of disk space. These formats are commonly addressed to as *scene-referred*. Several formats with possibly multiple encodings, listed in braces, are available: HDR (RGBE, XYZE), EXR (Half RGB) and TIFF (IEEE RGB, LogLuv24, LogLuv32), DICOM. The prominent HDR (.hdr) format allows to represent the scene with 76 orders of magnitude with

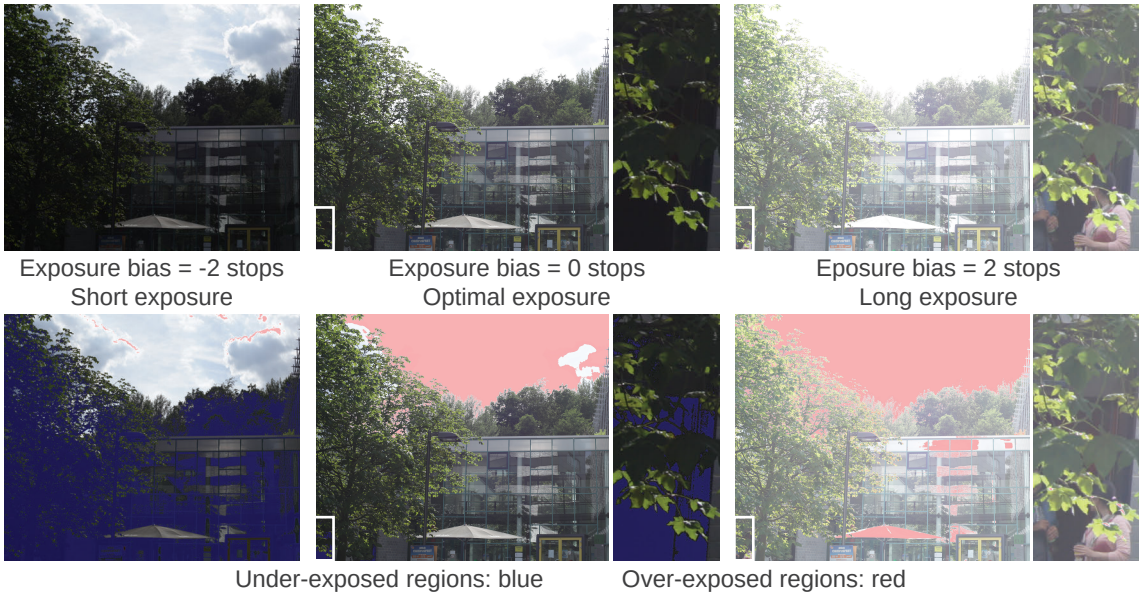


Figure 2.1: Example low dynamic range images of a scene with high dynamic range. The top row shows three different exposures of the same scene. In the bottom row, the over-/under-exposed regions in each of the shots are marked with red and blue, accordingly.

32-bit image and relative step size between adjacent quantized luminance values below 1 percent. Another popular format, the EXR (EXtended Range/.exr) format, is supported by an open-source C++ library, OpenEXR. It allows to represent the scene with 10.7 orders of magnitude with 48-bit image and relative quantization error up to 0.1 percent. HDR images and video can be stored not only in custom file formats, given above, but also in any standard compression format that supports higher bit-depths: e.g, MPEG4-AVC/H.264, JPEG2000 and JPEG XR. For more information on HDR images and video encoding as well as formats please refer to [Hoefflinger 2007, Reinhard et al. 2010, Mantiuk et al. 2015].

Since HDR images are scene-referred most image processing tasks benefit from HDR input compared to LDR input: This includes white-balancing, colour grading [Moroney et al. 2002, Reinhard et al. 2012, Fairchild 2013, Mantiuk et al. 2015], highlights editing and displaying [Mantiuk et al. 2015]. HDR images are used as light sources, representing the lighting effects which would be difficult to model in other way [Debevec 2008]. In satellite and medical imagery HDR content enables separate analysis of different wavelengths.

2.2 Dynamic Range Limitations in Conventional Capturing Systems

In a digital camera the input light is recorded with an image sensor, an electrical device which converts an input signal to an electronic signal. At a point in a sensor the light is integrated over the solid angle subtended by an aperture. Thus, the sensor records *irradiance*, defined as a radiant flux per unit area. Irradiance is measured in watt per square metre $[W/m^2]$. The irradiance is integrated over the time the shutter is open: The resulting quantity is called *exposure*, measured in lux-seconds [lx-s]. The exposure integrated over the pixel's footprint is then amplified and converted with an analog to digital converter (ADC) to an electronic signal.

There are two most spread types of sensors: CCD (charge-coupled device) and CMOS (complemen-

tary metal-oxide semiconductor). CCD sensors for long were considered as superior, meanwhile, due to recent progress in integrated circuit (IC) technologies, the quality obtained with CMOS sensors outperforms the one produced with CCDs. Moreover, CMOS sensors require less power, are better suited to high-speed capture and are cheaper to manufacture.

The sensors vary in size and resolution, as well as the size of each individual pixel. The size (namely its area) of each individual pixel influences dynamic range, which can be captured, and colour accuracy. The larger the pixels are the higher is a saturation threshold. Nevertheless, the low light conditions performance is not affected by the pixel size.

In order to measure the dynamic range captured with a digital camera the *signal to noise ratio (SNR)* measure is commonly used. Note that there exist multiple measures of dynamic range tailored to different applications [Reinhard et al. 2010, Mantiuk et al. 2015]. SNR is defined as the ratio of signal power to the noise power. It is measured in decibels [dB], which is ten times a base-10 logarithm of the ratio of two power quantities. One way to measure the SNR of a captured signal in a digital camera is to compute the ratio of signal and noise variance values. Often, the SNR is measured as the ratio of the intensity that starts to saturate the image sensor to the minimum intensity that can be observed above the noise level of the sensor. In this case, to convert to decibels, a base-10 logarithm of the ratio has to be multiplied by 20, since the power is proportional to the square of a signal.

The upper limit on the captured dynamic range is determined by a voltage well of a pixel. The low light performance is restricted by noise, that has a number of sources. The first source is the *dark current*, which is independent of light arriving at a sensor. Dark current is caused by the thermal random generation of electrons and holes. The next source is the *photon shot* noise. It is caused by the discrete nature of light and is the variance of the number of photons arriving at the sensor during the exposure time t. Finally, even if a sensor can capture large dynamic range, its usability is limited by the *readout noise*, which constitutes the precision at which light measurements gathered at the sensor are translated into digital values by an ADC. In order to process the input voltage a camera must amplify the pixel voltage to a range appropriate to an ADC. In high-level cameras only an additional uncorrelated noise is introduced during this step. This random noise can be reduced or eliminated by combining frames.

The output of an ADC is what is called a RAW camera format. Most digital cameras nowadays use 10 to 14-bit ADCs. Due to the limitations of displayable range the RAW format is typically compressed to an 8-bit format discussed in Section 2.1. In order to use the limited bit range and minimize the perceived noise, the gamma correction is typically applied to the RAW output of an ADC. The reason for this gamma correction comes from studies on perceived lightness: For perceptually uniform signal encoding it is sufficient to maintain a ratio of luminance values between adjacent codes of about a 1 percent. This can be achieved by coding the signal using a power law with an exponent between $1/3$ and 0.45 [Poynton 1996].

The RAW image format, supported by most of conventional cameras, is the closest to the high dynamic range image formats.

2.3 HDR Imaging Acquisition

In order to capture the full dynamic range of a scene various solutions are suggested: That might be hardware solutions, such as specialised optical systems and sensors (Section 2.3.1), or purely algorithmic solutions (Sections 2.3.2 and 2.3.3) targeted to conventional cameras.

2.3.1 Specialised Cameras

There exist specialised professional HDR video cameras (e.g., Arri Alexa XT and Red Epic Dragon) and custom-built ones [Nayar and Branzoi 2003, Tocci et al. 2011, Wan et al. 2012, Kronander et al. 2013, Manakov et al. 2013, Froehlich et al. 2014], providing high-end solutions that are mostly free of artefacts. Nonetheless, high cost, a significant level of required expertise as well as the need to manufacture a new camera hardware currently limit their use and availability.

The existing camera systems can be divided into two main groups. The first group comprises the approaches which modify how the sensor measures the light. The second group encompasses the optical systems, which use prisms, beam splitters, or mirrors to direct light to multiple sensors at different exposures.

I. Sensors Architectures There are multiple solutions to change the way the sensor measures the light. Probably the most simple approach is to use a logarithmic response for the sensor [Seger et al. 1999]. There are several commercially available cameras that use this technology, e.g., Viper FilmStream by Grass Valley. The dynamic range obtained with this technology is limited by the noise in dark scene regions.

Further, there exist solutions that allow varying the exposure time per pixel depending on the intensity of incident light [Brajovic and Kanade 1996, Ginosar and Gnusin 1997, Brajovic et al. 1998]. One of the first solutions is a so called 'sorting sensor' proposed by Brajovic and Kanade [1996]: It sorts pixels of sensed images by their intensities. In this architecture each sensing element is connected to a local processor. Each of these local processors in turn is connected to a global processor. The local processor is activated depending on the intensity received by a particular sensing element. Since the stronger singles elicit responses before the weaker one, the local processors are activated at different moments in time, and global processor works only with a few inputs at a time moment. The time the local processor triggers is proportional to the input radiation received. Further, while summing up the signals from all the local processors triggered at a given time moment, the global processor knows a number of the local processors, thus can build a cumulative histogram of the input intensity values. The sorting sensor handles dynamic range of approximately 6 orders of magnitude. Further, the strategy of exposing individual pixels or groups of pixels differently were used in several commercially available cameras. For example, in a commercial camera PIXIM, pixels are grouped with independent ADCs, then the conversion between the analog and digital signals stops either due to the set exposure time or due to the pixels values approaching saturation state. Alternatively, another commercially available camera SMaL (Camera Technologies of Cambridge) adjust pixels sensitivity values, instead of exposure times.

In order to achieve spatially varying exposures, a mask, which has a per pixel variable optical density, can be used. The mask can be either with fixed density values [Nayar and Mitsunaga 2000] or the optical density can adaptively vary [Nayar and Branzoi 2003, Nayar et al. 2004]. Nayar et al. [2006] suggested to use a programmable array of micro-mirrors to extend dynamic range. They experimented with several setups. The first is similar to the one suggested by Nayar et al. [2000] and uses four different exposure settings. They show that changing the relative positions of four exposures over time leads to less perceivable artefacts in videos compared to static images. The second setup implements the adaptive system introduced by Nayar and Branzoi [2003], but solves the limitations of their initial prototype: the blurring/diffraction effects are minimized and the optical efficiency is closer to 90%.

Hirakawa and Simon [2011] exploit different sensitivities in a regular Bayer pattern. Further,

Schöberl et al. [2012] proposed a method for HDR video capture based on spatially varying exposures, where individual pixels are covered with filters for different optical attenuation. The non-regular (random) pattern is used to minimize the aliasing. Then, Aguerrebere et al. [2014] proposed a new reconstruction method from a single shot acquired with a random pattern. This is a patch-based approach, extending the work of Yu et al. [2012] by accounting additionally for a noise level. The recent work by Serrano et al. [2016] further improves the reconstruction for high contrast scenes by exploiting recent advance in convolutional sparse coding.

Novel sensors such as the Aptina AR1331CP and Sony IMX135, increase the captured dynamic range by varying the electronic shutter time between pairs of alternate Bayer rows. For such specialized sensor architectures, the burden is on advanced off-line processing to avoid artefacts [Gu et al. 2010, Heide et al. 2014, Cho et al. 2014] due to rolling shutter distortion, motion blur as well as over- and under-exposure in interlaced pixel scanlines.

Dual ISO module on Canon cameras allows to set varying ISO setting for alternate Bayer rows [A1EX 2013]. The post-processing [A1EX 2013, Hajisharif et al. 2015] is required to avoid artefacts due to saturation and noise.

II. Optical systems There exist solutions that utilize beam splitters, which direct light to multiple sensors [Aggarwal and Ahuja 2004, Tocci et al. 2011, Kronander et al. 2013]. These solutions allow to avoid artefacts due to motion. The number of sensors is typically limited to 3-4, since any additional sensor complicates the light splitting optics, increases the cost of the camera and reduces the amount of light per sensor.

2.3.2 Multi-Exposure Techniques for Images

The simplest approach to capture extended dynamic range of a static scene with a conventional digital camera is to combine multiple images of the same scene, captured with different exposure times. There are two classes of multi-exposure algorithms: those that operate in the radiance domain and those that fuse differently exposed images in the image domain. Below we discuss each of two groups in more detail:

HDR Imaging Assuming the linear relation between the recorded values and the scene values (for example, this is the case for the images in a RAW format in non-saturated regions), the captured exposure values can be transferred to irradiance values by dividing each pixel value by an exposure time. In general case one has to account for a nonlinear response of a camera (e.g., gamma correction). The overview of camera response curve estimation strategies can be found in [Reinhard et al. 2010]. To obtain an HDR image, the corresponding pixels across multiple exposures are averaged, excluding overexposed or underexposed pixels. Prior to pixel averaging the images have to be aligned; as well as pixel's weight has to be selected carefully in order to avoid reconstruction artefacts due to camera and/or scene motion. For the detailed overview of exposure registration in the presence of motion see recent surveys [Srikantha and Sidibé 2012, Tursun et al. 2015].

Exposure Fusion Alternatively, the fusion can be performed in the image domain (e.g., [Mertens et al. 2007]). In that case the camera response curve is not reconstructed. Instead an LDR image is directly computed, that keeps only the best exposed parts of each of the exposures. This type of techniques corresponds to an "HDR mode" on mobile devices. Note, that exposure fusion algorithms do not produce actual high dynamic range images. Further, in order to work on a mobile device in real time these solutions should possess a certain level of computational efficiency, which

often comes at the cost of low accuracy results.

Independently of a fusion domain, an interesting problem is how to select exposure times. Here the target solution might address different goals, such as full dynamic range capture, acquisition of the minimal number of exposures or facilitation of images alignment.

Exposure Metering Multi-exposure HDR acquisition typically uses exponentially spaced exposure times [Mann and Picard 1995, Debevec and Malik 1997, Mitsunaga and Nayar 1999, Robertson et al. 2003], usually with a fixed ISO setting. One could also tabulate exposure times [Grossberg and Nayar 2003]. However, such approaches may use more exposures than necessary, especially if the scene histogram is sparse.

Spacing the exposures according to a Fibonacci sequence would be more efficient [Gupta et al. 2013], offering additional benefits for exposure registration. This technique is though targeted for professional high frame rate cameras. Other refinements are to incorporate noise models [Hirakawa 2010] or to vary ISO settings [Hasinoff et al. 2010]; the latter based on the observation that high ISO settings offer better signal-to-noise ratios.

Alternatively, metering algorithms can be applied so that a minimal number of exposures are selected to capture the scene with an optimal signal-to-noise-ratio [Barakat et al. 2008, Granados et al. 2010, Gallo et al. 2012].

2.3.3 Multi-Exposure Techniques for Videos

The multi-exposure techniques discussed in the previous section can be applied to HDR video capture. The main challenge in the reconstruction of HDR frames from LDR frames compared to the single images reconstruction lies in the importance of reconstruction not just a plausible image, but an image representing a specified moment of time.

Only the few techniques designed for single images allow the reconstruction of an HDR image with respect to a selected reference LDR frame [Sen et al. 2012, Hu et al. 2013, Bouzaraa et al. 2016]. Here, only well-exposed pixels from a reference exposure are used, while the remaining pixels are synthesized through analysis of similarity to corresponding regions in exposures captured before and after the reference exposure.

The existing work on HDR video capture is mainly focused on registration problem and selection of weighting strategy, while relatively no work exists on exposures selection for HDR video capture on standard cameras. Most of the exposure metering techniques discussed in Section 2.3.2 are designed to optimize between a large number of possible exposures and require collecting multiple images/histograms of a scene to recover its irradiance distribution. As such, they are not suitable for HDR video capture. One exception is a metering algorithm designed to find exposure parameters from a single frame [Ilstrup and Manduchi 2010], fitting a log-normal distribution to well-exposed pixels to estimate the scene's histogram. From this curve one can estimate the optimal exposure for the next frame.

To capture HDR video with multi-exposure technique, one could use costly high frame-rate cameras [Guthier et al. 2013], or advanced post-processing could be applied [Kang et al. 2003, Sen et al. 2012, Kalantari et al. 2013]. Such algorithms synthesize the missing luminance information for each frame, often with the aid of optical flow techniques. Here, motion estimation is particularly complicated as pixel differences are due to a combination of motion and exposure time differences. This may complicate the detection of pixel correspondences and cause ghosting artefacts [Srikantha

and Sidibé 2012]. Namely, this is often the case when objects exhibit non-rigid motion [Kalantari et al. 2013, Fig. 6]. Additionally, such merging approaches are designed to require a regular and repeating sequence of exposures, and therefore cannot be directly combined with advanced metering techniques.

PatchMatch to a large extent allows to ameliorate the ill-effects of scene motion [Sen et al. 2012], allowing for HDR video reconstruction [Kalantari et al. 2013]. Moreover, this technique is amenable to reconstructing HDR video from sequences of more than 2 exposures. It is, however, relying on regular repeating patterns of exposure timing, precluding the use of metering strategies. Further, if over-exposed image areas do not follow the dominant motion between frames, this algorithm may still produce visible artefacts.

2.3.4 Summary

The advantage of the techniques which modify the sensor measurement strategy is that they allow to obtain an HDR frame in a single shot. It expels the need for motion estimation. Further, while the reconstruction of overexposed pixels is needed, these pixels are not grouped in large regions as in multi-exposure techniques, and thus the reconstruction can be performed more reliably.

At the same time, for static scenes the multi-exposure techniques are beneficial due to the fact that the information is contained at least in one of exposures and no reconstruction of missing information is needed. In case of video, the reconstruction problem is becoming underconstrained, especially in the presence of fast motion. In that case the special attention has to be designated to the exposure selection to prevent loss of information in too large regions. For video capturing, the better approach might be to use spatially varying coding on a sensor. The usage of these approaches is currently limited by the cost and the requirement to manufacture a new hardware.

2.4 Calibration

To calibrate images to accurate radiometric values, they need to be linearised first, and then scaled. Although linearisation of images has received considerable attention, algorithms for estimating absolute luminance values in images are few and far between, in particular those that do not require extraneous equipment. Note that HDR images obtained with multi exposure techniques are linear by construction. An image can be calibrated to radiometric values either by performing direct measurements of the scene (Section 2.4.1), an initial camera calibration based on scene measurements (Section 2.4.2) or by relying on EXIF data, such as exposure time, ISO value and aperture size (Section 2.4.3).

2.4.1 Scene Measurements

Conventionally, to obtain precise absolute values from an image, an 18% grey card or a colour checker is inserted into a scene. Then, the reference colour (colours) is measured by a photometer/spot luminance meter.

2.4.2 Camera Characterization

Another strategy to get images in absolute values is to perform an initial camera calibration [Wüller and Gabele 2007, Fliegel and Havlin 2009], which is valid only for a particular combination of a camera and a lens, as well as camera settings.

The approach estimates an OECF (opto-electronic conversion function) of a camera. The OECF is a relationship between scene luminance and digital output levels of an opto-electronic digital image capture system. The initial calibration requires to perform the scene measurements as described in Section 2.4.1. Then, in order to use a camera as a luminance meter one needs to know the OECF as well as the exposure value of calibration.

Not all cameras though can be used: It is important to test that characteristic response of a digital camera remains unchanged for different exposure times. The image-processing algorithms performed by a camera, details of which are usually not provided by camera manufacturers, lead to errors in measurements results.

2.4.3 EXIF Based Approach

Under some assumptions on the camera properties, such as lens transmittance and vignetting factor values, used definition of sensitivity and camera metering mode one could get an approximation of light conditions. In this section we describe different approaches for absolute values approximation from EXIF data.

The incoming scene radiance L is transformed by an optical system into the focal plane exposure H which is proportional to the integration time t, but decreases with the f-number N of a lens

$$H = Et = \left(q\frac{L}{N^2}\right)t, \tag{2.1}$$

where E is the irradiance which arrives to a camera sensor and q is the proportionality factor which could be derived as follows.

The irradiance at the pixel with the area A_p can be found as

$$E = \frac{d\Phi_i(A_p)}{dA_p}, \tag{2.2}$$

where $\Phi_i(A_p)$ is the radiant power that arrives at the pixel with the area A_p from the solid angle subtended by the lens as seen from the pixel. It can be written as

$$\Phi_i(A_p) = \int_{A_p}\left(\int_{\Omega_p^j} L_i(p^j,\omega_i^k)d\sigma^{\perp}_{p^j}(\omega_i^k)\right)d\mu^2(p^j), \tag{2.3}$$

where Ω_p^j is the solid angle subtended by the lens as seen from $p^j \in A_p$, $\sigma^{\perp}_{p^j}(\omega_i^k)$ is the projected solid angle measure, $\omega_i^k \in \Omega_p^j$, $d\mu^2(p^j)$ is the differential measure on A_p and $L_i(p^j,\omega_i^k)$ is the radiance arriving at p^j from the direction $\omega_i^k \in \Omega_p^j$ (Figure 2.2).

Let assume that radiant power flows lossless through the lens: Then, the flux $\Phi_i(A_p)$ will be equal to the flux $\Phi_o(A_s)$ leaving the projection A_s of the pixel on the object's surface in the solid angle subtended by the lens as seen from A_s:

$$\Phi_i(A_s) = \int_{A_s}\left(\int_{\Omega_s^j} L_o(s^j,\omega_o^k)d\sigma^{\perp}_{s^j}(\omega_o^k)\right)d\mu^2(s^j), \tag{2.4}$$

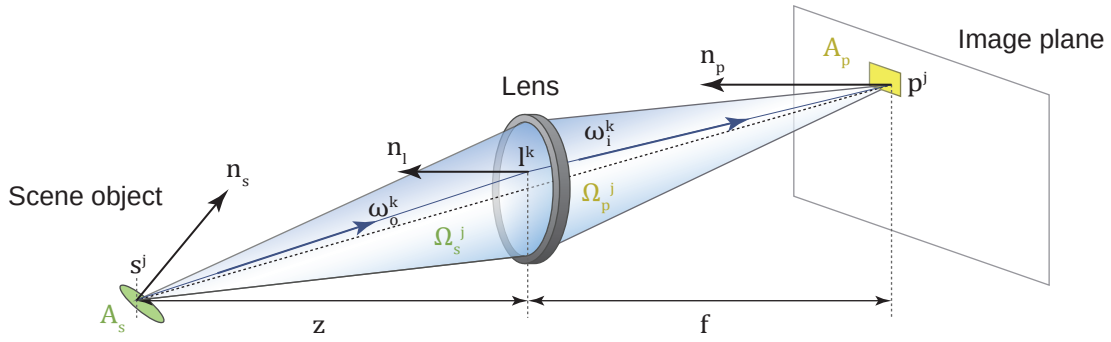


Figure 2.2: Irradiance on the image plane. Please refer to the main text for the explanation of the used symbols.

where Ω_s^j is the solid angle subtended by the lens as seen from $s^j \in A_s$, $\sigma_{s^j}^{\perp}(\omega_o^k)$ is the projected solid angle measure, $\omega_o^k \in \Omega_s^j$, $d\mu^2(s^j)$ is the differential measure on A_s and $L_o(s^j, \omega_i^k)$ is the radiance leaving s^j in the direction $\omega_o^k \in \Omega_s^j$.

Under the assumption that the radiance distribution is constant in the scene $L_o(s^j, \omega_o^k) = L_o = const$, it can be moved outside the integral:

$$\begin{aligned} \Phi_i(A_p) &= \Phi_o(A_s) = \\ &= L_o \int_{A_s} \left(\int_{\Omega_s^j} d\sigma_{s^j}^{\perp}(\omega_o^k) \right) d\mu^2(s^j). \end{aligned} \tag{2.5}$$

This equation can be simplified, assuming that the solid angle Ω_s^j is the same for each point $s^j \in A_s$: $\Omega_s^j = \Omega_s$ for each s^j. Then the two integrals comprising the double integral are independent of each other:

$$\begin{aligned} \Phi_i(A_p) &\approx L_o \int_{A_s} \left(\int_{\Omega_s} d\sigma_s^{\perp}(\omega_o^k) \right) d\mu^2(s^j) = \\ &= L_o \int_{A_s} d\mu^2(s^j) \int_{\Omega_s} d\sigma_s^{\perp}(\omega_o^k) = \\ &= A_s \int_{\Omega_s} d\sigma_s^{\perp}(\omega_o^k). \end{aligned} \tag{2.6}$$

The integral $\int_{\Omega_s} d\sigma_s^{\perp}(\omega_o^k)$ is called the projected solid angle associated with solid angle Ω_s.

The projected solid angle measure $d\sigma_s^{\perp}(\omega_o^k)$ can be transformed into the Lebesgue area measure as follows:

$$\begin{aligned} d\sigma_s^{\perp}(\omega_o^k) &= |< n_s, \omega_o^k >| d\sigma_s(\omega_o^k) = |< n_s, \omega_o^k >| \frac{|< n_l, \omega_o^k >| d\mu^2(l^k)}{||s - l^k||^2} \\ &= \{||s - l^k||^2 = z^2 / |< n_l, \omega_o^k >|^2\} = \\ &= |< n_s, \omega_o^k >| \frac{|< n_l, \omega_o^k >|^3 d\mu^2(l^k)}{z^2} = \\ &= \{\cos\theta^k = |< n_s, \omega_o^k >| \quad \text{and} \quad \cos\alpha^k = |< n_l, \omega_o^k >|\} = \\ &= \cos\theta^k \frac{\cos(\alpha^k)^3 d\mu^2(l^k)}{z^2}. \end{aligned} \tag{2.7}$$

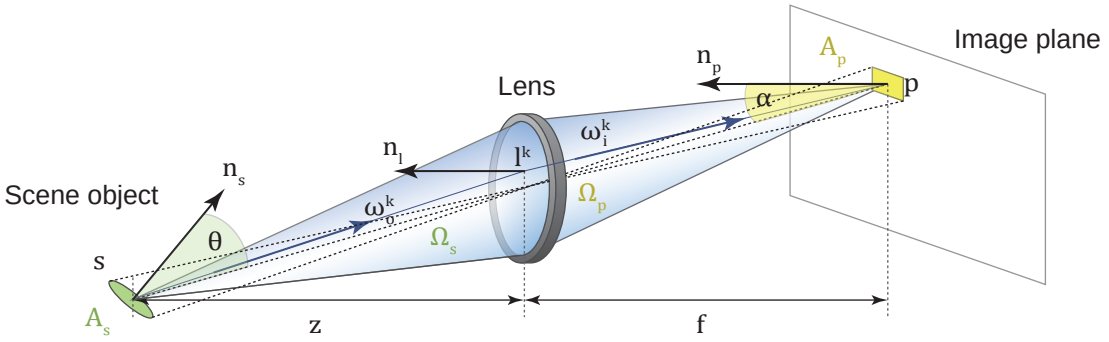


Figure 2.3: Simplification of a camera model to a pinhole camera model. Please refer to the main text for the explanation of the used symbols.

Further, assuming a pinhole camera model, $\cos\theta^k$ can be assumed equal to $\cos\theta$, where θ is the angle between the line, passing through the lens centre and connecting the pixel with its projection to the object, and the normal to the area A_s (Figure 2.3); as well as $\cos\alpha^k$ can be assumed equal to $\cos\alpha$, where α is the angle between the line, passing through the lens centre and connecting the pixel with its projection to the object, and the normal to the image plane (Figure 2.3). Then, Equation 2.6 can be simplifies as follows:

$$\begin{aligned}\Phi_i(A_p) &= A_s \int_{\Omega_s} d\sigma_s^{\perp}(\omega_o^k) \\ &\approx A_s \cos\theta \frac{\cos(\alpha)^3}{z^2} \int_{A_l} d\mu^2(l^k) \\ &= A_s \cos\theta \frac{\cos(\alpha)^3}{z^2} \frac{\pi d^2}{4}\end{aligned} \tag{2.8}$$

where l^k is a point on the lens, $d\mu^2(l^k)$ is the differential measure on the lens with the area A_l and d is the diameter of the lens.

Thus, assuming that the radiant power flows lossless through lens, by substituting Equation 2.8 into Equation 2.2 we get:

$$E = L \frac{dA_s}{dA_p} \cos\theta \frac{\cos(\alpha)^3}{z^2} \frac{\pi d^2}{4} \tag{2.9}$$

In general case this equation also includes the transmittance T of the lens and the vignetting factor $v(\alpha)$:

$$E = \frac{d\Phi_i(A_p)}{dA_i} = T v(\alpha) L \frac{dA_s}{dA_p} \cos\theta \frac{\cos(\alpha)^3}{z^2} \frac{\pi d^2}{4}. \tag{2.10}$$

Using equality of the solid angles $d\omega_s$, subtended by the area A_s as seen from the lens centre, and $d\omega_p$, subtended by the area A_p as seen from the lens centre:

$$\frac{dA_s \cos\alpha}{(z/\cos\theta)^2} = \frac{dA_p \cos\alpha}{(f/\cos\alpha)^2}, \tag{2.11}$$

we get:

$$\frac{dA_s}{dA_i} = \frac{\cos\alpha}{\cos\theta} \left(\frac{z}{f}\right)^2. \tag{2.12}$$

From this and the definition of the f-number $N = f/d$ the final relation between L and E is

$$E = Tv(\alpha)L\frac{(\cos\alpha)^4}{N^2}\frac{\pi}{4}. \tag{2.13}$$

Generally it is assumed that $T = 0.9$, $\alpha \in [6^o\, 10^o]$ and $v = 0.98$ and thus

$$q = \frac{\pi}{4}Tv(\alpha)(\cos\alpha)^4 \in [0.65\, 0.6777]. \tag{2.14}$$

When the exposure value H is chosen automatically by a camera its value depends on used definition of sensitivity as well as the camera metering mode.

Exposure Based on Sensitivity S

ISO standard 12232 [ISO 2006] includes several definitions of sensitivity. Usually it is not apparent from the available meta-data which definition was used. This can lead to greatly differing exposures for the same pixel value [Ackermann and Goesele 2013]. The two most common definitions of sensitivity are Saturation-based Sensitivity and Standard Output Sensitivity.

Sensitivity could be defined as:

$$S = \frac{10}{H} = \frac{10}{Et}. \tag{2.15}$$

From (2.13) and (2.14) it can be derived:

$$S = \quad 14.76\frac{N^2}{Lt}, \quad for\,\alpha = 6^o \tag{2.16}$$

$$S = \quad 15.4\frac{N^2}{Lt}, \quad for\,\alpha = 10^o \tag{2.17}$$

In general form, the equation connecting the mean luminance L and the sensitivity S has the following form:

$$S = K\frac{N^2}{Lt}. \tag{2.18}$$

Standard Output Sensitivity S_{SOS} This definition assumes that the mean luminance value of a scene corresponds to a luminance value of a standard gray card with 18% reflectivity (r = 0.18), if this card was present in the scene. Then, the mean luminance of a scene is mapped to an intensity value $Z_{0.18} = 0.18^{(1/\gamma)}$, where γ is a display gamma defined by a colour space. The image values are normalised to a range from 0 to 1.

Many standard colour spaces such as sRGB and Adobe RGB are designed for display with $\gamma = 2.2$. In this case the normalized pixel level is 0.4586, what is the same as the pixel level equal to 116 for 8-bit pixels where the maximum is 255. The value 0.461 or the pixel level 118 is used in DC-004 because sRGB gamma is not exactly 2.2. The standard output sensitivity is defined as:

$$S_{SOS} = K\frac{N^2}{L_{0.18}t}, \tag{2.19}$$

where $L_{0.18}$ is the absolute mean luminance of a grey card.

In order to computer a scale factor, one has to reverse an Equation 2.19 to compute $L_{0.18}$ and divide it by $Z_{0.18}$.

Saturation-based ISO sensitivity S_{sat} This definition as well assumes the mean luminance value of a scene corresponds to a luminance value of a standard gray card with 18% reflectivity (r = 0.18), if this card was present in the scene. The difference is that now, the sensor saturates (reaches its maximum output) at 141% reflectivity in order to provide some "headroom".

The equation connecting the average scene luminance L and the sensitivity S is derived from the assumption that the radiance of an 18% gray card $L_{0.18}$ is 18/141 of the value that saturates a sensor, L_{sat}:

$$L_{0.18} = (18/141.4)L_{sat} = 0.1276L_{sat}. \tag{2.20}$$

Using the form (2.16) of an equation (2.13) one can get:

$$S_{sat} \quad = \quad K\frac{N^2}{Lt} = K\frac{N^2}{0.1276L_{sat}t}. \tag{2.21}$$

Given the values of S_{sat}, N and t the scale factor can be calculated by inverting an Equation 2.21 to estimate L_{sat}. Then, the corresponding intensity value Z_{sat} is calculated by extrapolating the brightest unsaturated patches in the image, as the saturated patches do not contain information. Finally, the scale factor is just the ratio of L_{sat} and Z_{sat}.

[Ackermann and Goesele 2013] demonstrated that applying these two definitions of sensitivity to the same pixel gives two estimations, where one exceeds another by a factor of 1.43.

Exposure Based on the Reflected Light Meter Calibration Constant

The constant K in an Equation 2.18 is frequently called the in-camera reflected-light meter calibration constant and depends on a camera. The common choice of this constant is $K = 12.5$.

Absolute Data Estimation

A scale factor for an image can be obtained from the equations of the form (2.18) with an appropriate value of K and under assumption that the mean luminance in the area measured by in-camera light meter corresponds to the luminance of 18% card if this card was inserted into the scene and measured. Further, the accuracy of the estimate depends on the knowledge of the sensitivity definition which is used in a camera. The first assumption might lead to large errors in case of high dynamic range scenes depending on the metering mode.

Ackermann et al. [2013] used an equation 2.18 to estimate the absolute values for the images downloaded from the internet. The constant K was set to 15.4, standard output sensitivity definition was assumed. By explicitly accounting for a gamma curve in the sRGB colour space the absolute luminance values can be obtained as follows:

$$L_p = 15.4\frac{f^{-1}(Z_p/255)}{f^{-1}(118/255)}\frac{N^2}{S_{SOST}}, \tag{2.22}$$

where f^{-1} corresponds to the inverse gamma correction of sRGB colour space, L_p is an absolute luminance value in pixel p with intensity value Z_p.

2.4.4 Summary

In order to convert intensity values stored in an HDR image to their radiometric counterparts with high accuracy, one has to perform direct scene measurements. In case if measuring equipment is not accessible one can approximate a scaling factor from EXIF data. The computation of the scale factor for conversion to absolute luminance values requires the camera specification data. The accuracy depends on the metering mode as well as a scene. The initial camera characterization helps to improve the accuracy of estimation from EXIF data. There exists no solution that allows to obtain actual radiometric values from a single image without any additional information.

2.5 Light Fields

The *light field* is a general representation of the light traversing a scene, which provides information about its appearance observed from different directions. The term light field was introduced by Gershun [1936]. Its definition coincides with a definition of *plenoptic function* introduced by Adelson and Bergen [1991]: The radiance at a point in a given direction. The plenoptic function is a five dimensional quantity describing the flow of light at every 3D spatial position (x,y,z) for every 2D direction (θ,ϕ). It is sometimes given with the two additional dimensions: time and wavelength. As was observed simultaneously by [Levoy and Hanrahan 1996, Gortler et al. 1996] the 5D representation may be reduced to 4D in regions free of occluders. This is a consequence of the fact that in the absence of occluders light intensity does not change while travelling along a ray. A 4D light field can be considered as a collection of pinhole views, where cameras centres are lying in the same plane and are equidistant from each other. The distance between the cameras is often called light field *baseline*.

A common approach to parametrize a 4D light field L is to use a two-plane parametrization:

$$L : \Pi \times \Omega \to \mathbb{R}, \quad (s,t,x,y) \to L(s,t,x,y), \tag{2.23}$$

where (s,t) defines a view in the lens plane $\Pi \in \mathbb{R}^2$ and (x,y) is a spatial position within the image plane $\Omega \in \mathbb{R}^2$. This parametrization was used for example in pioneering works on light field [Levoy and Hanrahan 1996, Gortler et al. 1996]. We further will instead use the parametrization introduced by Bolles et al. [1987]. In this case the $(x,y) \in \mathbb{R}^2$ coordinate is defined as a spatial position within each of sub-aperture views individually in contrast to within a whole image plane $\Omega \in \mathbb{R}^2$. In other words, in case of the latter parametrization, the coordinate (x,y) is given relative to the coordinate (s,t). Both types of parametrization are equally likely to be found in the literature.

Given the completeness of light fields, they have a potential of becoming the ultimate visual data representation for digital imaging. Thus, representing scene as a light field potentially enables reproducing all visual cues that can be observed in the real world, including motion parallax and accommodation, which are usually missing in conventional imaging pipelines.

The light-field input simplifies image-based depth estimation. It as well enables a range of effects in post-production. Among the most common are a change of view, depth of field rendering and alpha-mating (e.g., [Levoy and Hanrahan 1996, Zilly et al. 2013]).

Further, the development of glasses-free auto-stereoscopic displays and light-field displays (e.g., [Hirsch et al. 2014, Efrat et al. 2016]) requires generation of content. A recent summary of 4D light-field capturing techniques can be found in a book chapter called "Plenoptic cameras" by Goldluecke et al. [2015].

Light-field data requires large storage space, thus the proper compression method is of high importance. For instance, Kim et al. [2013] suggest for each point in the scene to keep its estimated depth and colour value. Additionally, they store angularly varying information, such as the appearance of specular regions and occlusion boundaries.

2.6 Registration Problem in Light Fields

In stereo image pairs, depth is estimated by finding correspondences between the points of two views. The correspondences are defined by the displacement (disparity) of the same object point between the views. The disparity d and depth Z values are linearly related:

$$Z = bf/(b-d), \tag{2.24}$$

where f is a focal length and b is a baseline, both in pixels units.

In a structured light field L, the displacement between cameras is uniform, and therefore, a single disparity value d provides a correspondence between all instances of the same scene point. Each point in a 4D light field L can be represented by four coordinates (s,t,x,y), using two plane parametrization (2.23). In the absence of occlusions, the instance of the point $L(s,t,x,y)$ in other views can be obtained from the offset $(\Delta s, \Delta t)$ between the views:

$$L(s+\Delta s, t+\Delta t, x+\Delta sd, y+\Delta td) = L(s,t,x,y). \tag{2.25}$$

Each of the views in a light field is also often called a *sub-aperture*.

Since a 4D light field is a dense collection of multiple views, any conventional correspondence estimation method can be used. However, light fields contain rich information, which allows to develop more robust and efficient specialized methods. Thereupon, the most common approaches for depth estimation from light field are:

- estimation of oriented point cloud by applying patch-based feature matching;
- computing per-image depth maps using dense correspondence matching.

Feature matching approaches cannot match thin features and tend to produce errors in textureless regions. Since in our work we are interested in high quality depth reconstruction, we discuss in more details the second type of approaches, targeted to light field input. These approaches can be subdivided into two main groups: those which work on 2D slices of the light field and those that exploit the full angular information in the light field. Further, the approaches can be divided into those which use global regularization with certain smoothness assumptions and those which do not. For a recent detailed overview of multi-view stereo please refer to [Furukawa et al. 2015, Eisemann et al. 2015].

In Section 2.6.1 we first discuss the approaches that estimate depth from 2D slices of the light field. Then, in Section 2.6.2 we discuss the algorithmic solutions based on point consistency. Moreover, in Section 2.6.3 the depth estimation methods for non-lambertian objects are discussed separately. Finally, we conclude this chapter with a summary in Section 2.6.4.

2.6.1 Epipolar Plane Images Analysis

We start from discussion of the first group of approaches, which work with 2D slices and do not exploit the full information stored in a light field. The 2D slice of a light filed is obtained as a slice

of the light field along an epipolar line. The *epipolar line* is formed by crossing of an epipolar plane and two images. The *epipolar plane* is defined as follows: For each point P in the scene, there is a plane, called the *epipolar plane*, that passes through the point and the line joining the two lens centres. For a rectified light field epipolar lines are aligned with pixel scanlines. Then, an *Epipolar Plane Image (EPI)* is the 2D image in the plane of the cut through the stack of views along an epipolar line. In other words, it is the image formed by pixel scanlines in different sub-aperture views, lying on the same epipolar line, stacked on top of each other. An example EPI for a dense three dimensional light field, consisting of 100 horizontal views, from a dataset published by Kim et al. [2013], is given in Figure 2.4.



Figure 2.4: An example EPI for a dense light field from a dataset published by Kim et al. [2013]. The top row shows three views from the input light field with the following dimensions: $(s,t,x,y) = (100,1,4020,2679)$. The white dashed line denotes an epipolar line, crossing the light-field views in pixels with vertical coordiante $y = 1150$. These crossings stacked on the top of each other form the EPI image in the bottom. The images are stacked so that the first image is on the top and the last is on the bottom.

The algorithms which try to estimate depth based on the analysis of EPIs typically assume dense light field input and rely on the following properties of EPIs. A feature track of a diffuse point in a structured light field is a line, where the slope of the line uniquely defines the depth of the point. Moreover, the larger is the slope, the closer is the point to a camera (Figure 2.5a). Further, the feature track of the point can be occluded only by feature tracks with larger slopes (Figure 2.5b). The algorithms, then, rely on the assumption that feature track lines can be more robustly detected than point correspondences.

The basic insight to leverage these structures for scene reconstruction was proposed as early as 1987 [Bolles et al. 1987]. Criminisi et al. [2005] suggested to extract EPI tubes (a group of EPI lines with the same slope) from EPIs iteratively to account for occlusions. The tubes are extracted by analysing photo-consistency along a certain slope direction. In a work of Wanner and Goldluecke [2012], which is considered to be a canonical work on depth estimation from light field input, depth labelling problem is formulated as a constrained optimisation problem on EPIs. They first estimate local depth estimates by computing structure tensor, which allows to compute the direction of local lines as well as the reliability of the estimates. Then, in order to account for global consistency, such as ordering of lines with respect to occlusions in EPI, the appropriate constraints are formulated and a discrete optimization problem is solved. Structure tensor computation involves an initial smoothing that might lead to errors in edges location. To alleviate this problem during the final integration of depth estimates from horizontal and vertical EPIs, the additional term is considered in the optimization problem, which account for edge strength in EPI. In [Goldluecke and Wanner 2013] the global regularization is performed by formulating photo consistency requirement as an anisotropic total variation. The local constraints on depth ordering in EPIs caused by occlusions are formulated as a differential equation. This approach relies on the fact that the baseline is

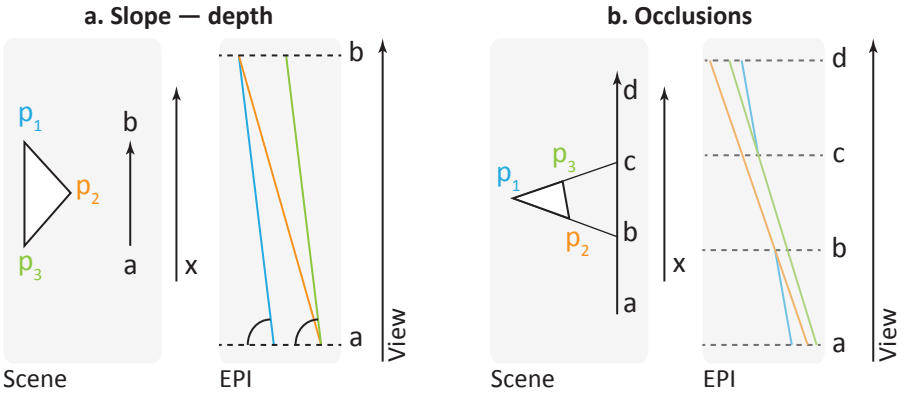


Figure 2.5: Example scenes and their corresponding EPI images adapted from [Bolles et al. 1987]. **(a)** The two points p_1 and p_3 are equally distant from a camera plane $a-b$. Then, the feature tracks of points p_1 and p_3 have the same slopes, denoted by equal angle between their feature tracks and horizontal axis. A point p_2 lies closer to a camera plane than points p_1 and p_3, and thus the slope of its feature track is larger than the slopes of feature tracks of points p_1 and p_3. **(b)** In some views the point can be occluded by the points, which are located closer to a camera plane. In this example the point p_1 is occludded in the views between points b and c.

small. Since this regularization scheme is very slow, in [Wanner and Goldluecke 2014] the authors experiment with simplified regularizing schemes, by simply enforcing the depth discontinuities to lie on edges of the input image. The horizontal and vertical estimates are integrated by choosing the one with higher reliability value. For the optimal performance this algorithm require around 9 views and disparity values smaller than 1 pixel. Kim et al. [2013] propose a fine-to-coarse approach that does not require global regularization. The algorithm is targeted to 3D light fields of very high spatio-angular resolution. The depth is estimated first on the highest resolution level and assigned only for the points that have confident depth estimates: this is the case for the points lying on the object edges. A distinctive feature of this algorithm is that depth estimates are propagated to all views. The disadvantage of this algorithm is that the returned depth values are discrete, and there is no way in advance to select a good range of these disparities values. Further, Yucer et al. [2016] recently presented an extension of these ideas to the case of unstructured dense light field, obtained as a video. Depth is estimated locally, on EPIs constructed for image pairs, by exploiting gradient directions. As in previous work, the depth is first computed only for high-gradient regions. Then, depth is propagated within each of the images. The proposed consistency measure is used to determine if the edge is due to a texture or due to occlusions, based on what a depth propagation direction is selected. The depth maps are then aggregated into a voxelized 3D scene with consequent refinement to an oriented point cloud.

2.6.2 Full 4D Angular Consistency

Alternatively a photo-consistency measure can be computed on the angular sampling image [Wood et al. 2000, Yu et al. 2004], constructed from the whole light field. This image typically is referred to as Surface Light Field [Wood et al. 2000] or *Surface Camera (SCam)* [Yu et al. 2004]. Given a point in a light field L with coordinates (s,t,x,y) and disparity value d, the corresponding SCam $S_{s,t,x,y,d}$ can be constructed, based on Equation (2.25), as a union of all rays traced from a corresponding

scene point to each of the views:

$$S_{s,t,x,y,d} = \bigcup_{(\Delta s,\Delta t)} L(s+\Delta s, t+\Delta t, x+\Delta sd, y+\Delta td). \tag{2.26}$$

Examples of 1D and 2D SCams are given in Figures 2.6 and 2.7. Note, that in presence of occlusions the pixels of SCam can correspond to different scene points (Figures 2.6a, 2.7c).

In absence of occlusions and under an assumption of a diffuse Lambertian object (e.g., Figure 2.6), the photo consistency measure computed on such angular sampling image will have the highest value, when the SCam is computed with ground-truth disparity value. Without explicit occlusion consideration the resulting depth map will have smooth transitions on object edges, leading to misalignment of edges.

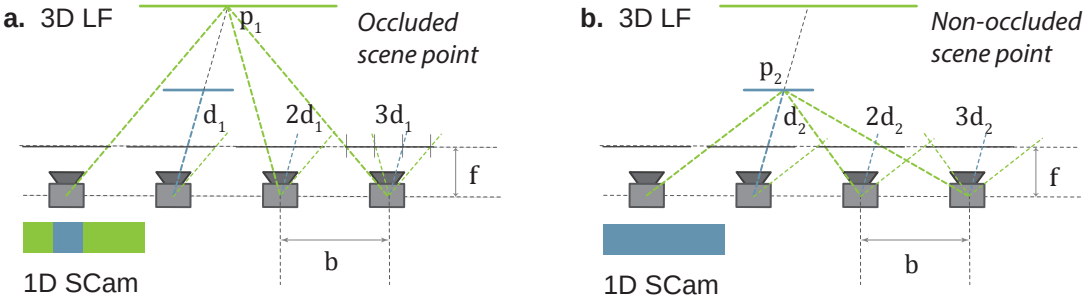


Figure 2.6: An illustration of a 1D SCam. The 3D light field consists of two planar objects (green and blue) at two different depths. A 1D SCam is shown for the scene point where the dashed lines (rays) converge, as the image of the point in the different views.

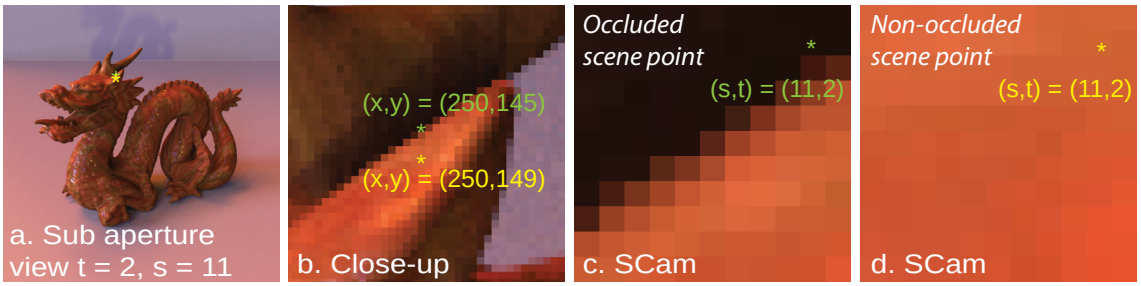
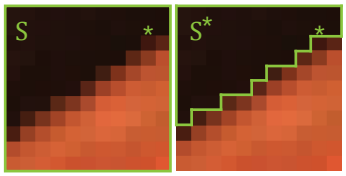


Figure 2.7: Illustration of a 2D SCam. **(a)** A view of a light field with coordinates $(t,s) = (2,11)$. **(b)** A close-up of this view. **(c)** An SCam corresponding to a scene point with coordinates $(y,x) = (145,250)$ in a view with coordiantes $(t,s) = (2,11)$. This point is indicated with a green star. **(d)** An SCam corresponding to a scene point with coordinates $(y,x) = (149,250)$ in a view with coordiantes $(t,s) = (2,11)$. This point is marked with a yellow star.



In presence of occlusions the SCam might be divided into several regions: Thus, one possible solution is to find a set of pixels $S^*_{s,t,x,y,d} \subseteq S_{s,t,x,y,d}$, which are the actual instances of the scene point observed in a light-field point with coordinates (s,t,x,y), prior to photo-consistency computation. An example S^* for an SCam image in Figure 2.7c is given on the left.

Chen et al. [Chen et al. 2014] determine $S^*_{s,t,x,y,d}$ as the pixels with the most similar colour values to a colour of reference pixel with coordinates (s,t,x,y). To this end, a bilateral consistency metric

(BCM) for each SCam pixel is computed with respect to a reference pixel based on spatial and colour distances. Then, $S^*_{s,t,x,y,d}$ is obtained by thresholding the values of BCM. Finally, for each test value $\tilde{d}$ the photo consistency measure $\rho_{\tilde{d}}(s,t,x,y)$ is computed on the region $S^*_{s,t,x,y,\tilde{d}}$. The photo consistency measure is defined as an exponential function on colour distances, which is shown to outperform a standard l_2 norm. The final disparity value for a considered light-field pixel is estimated as the one that gives the optimal value of the consistency measure: $\mathbf{d}(s,t,x,y) = \underset{\tilde{d}}{\operatorname{argmax}}\, \rho_{\tilde{d}}(s,t,x,y)$.

It is further shown that when the occluder has a similar colour to a reference point, a correct depth might not correspond to a global minimum of the consistency measure, but it always corresponds to one of local minimums. Moreover, for non-textured regions there can be a valley of minimum values. In order to avoid erroneous estimates in such regions the spatial neighbourhood is considered and the depth estimate confidence is weighted based on how textured is a region in a sub-aperture view. Finally, only the pixels, for which there is a significant gap between the global minima and the second smallest local minima of the weighted consistency measure, are marked as reliable and assigned the depth corresponding to a global minima of the consistency measure. For the rest, the depth labelling is obtained through a within sub-aperture view depth propagation. This approach produces high quality depth estimates for noise-free images.

Wang et al. [2015] observe that the occluder edge orientation can be predicted by the edge orientation in the spatial domain (for instance, compare the edges in SCam in Figure 2.7c and in the sub-aperture view in Figure 2.7b). The disadvantage of this approach is that it can not handle small details and the final result is highly dependent on edge estimation step. Sheng et al. [2015] as well exploit the fact that the occlusion shape in SCam is similar to the one in the image: A guided integral filter is built to predict the occlusions. The consistency measure is weighted by the pixel probability of being an occluder.

A different approach is suggested by [Lin et al. 2015]. They model the depth estimation as an energy minimization problem. Their consistency measure consists of three terms:

- C1 The standard consistency measure across all sub-aperture views, computed on colours and gradients, does not account for occlusions.
- C2 The consistency measure defined on a focal stack: *The focal stack is a sequence of images captured with different focus settings. Each image in the stack is obtained from a light field with the focal value proportional to one of trial disparity values. The number of images in the stack is equal to a number of considered disparity labels. Then, for each point in the central sub-aperture view the sum of differences between the pixel value in the focal stack slice and the approximation of the same value from the central view is computed over a small neighbourhood of the current estimate of a disparity value.*
- C3 The consistency measure based on the observation of local symmetry properties of the points in the slices of the focal stack: *The main observation is that if a current focal value corresponds to an actual depth of the considered point, then in the absence of occlusions the colour values will be locally symmetric. They formulate two alternative measures for local symmetry depending on the presence or absence of occlusions. An occlusion probability map is constructed by combining all three consistency measures computed under the assumption of no occlusions.*

This global approach is robust to noise and tend to produce good depth estimation results. Williem and Kyu Park [2016] show that the consistency measure $C2$ taken alone performs worse than other consistency measures.

Kamal et al. [2015] formulate a minimization problem in local patches across views with a careful discretization of disparity values. The Lambertian surface is assumed and occlusion, inter reflection and specularities are modelled as a sparse error term.

Williem and Kyu Park [2016] perform a comparison of several photo consistency measures and suggest the new one robust to occlusions and noise simultaneously. They show that for noise free results their and Chen et al. [2014] methods perform the best. Chen et al. [2014] approach performs even slightly better in absence of noise, while fails drastically in case of noisy data. The quality of new proposed method stays comparable to noise-free performance even for very noisy input.

Finally, Heber and Pock [2016] utilize a convolutional neural networks to predict the orientation of feature track lines in EPIs. They further use 4D regularization step to propagate the information to textureless regions.

2.6.3 Depth Estimation for Non-Lambertian Surfaces

In the presence of angularly varying effects such as specular highlights, the assumption about angular colour consistency for scene point does not hold anymore. The SCams of a specular point will have smooth appearance. In EPIs the feature tracks of specular features lie on the top of diffuse

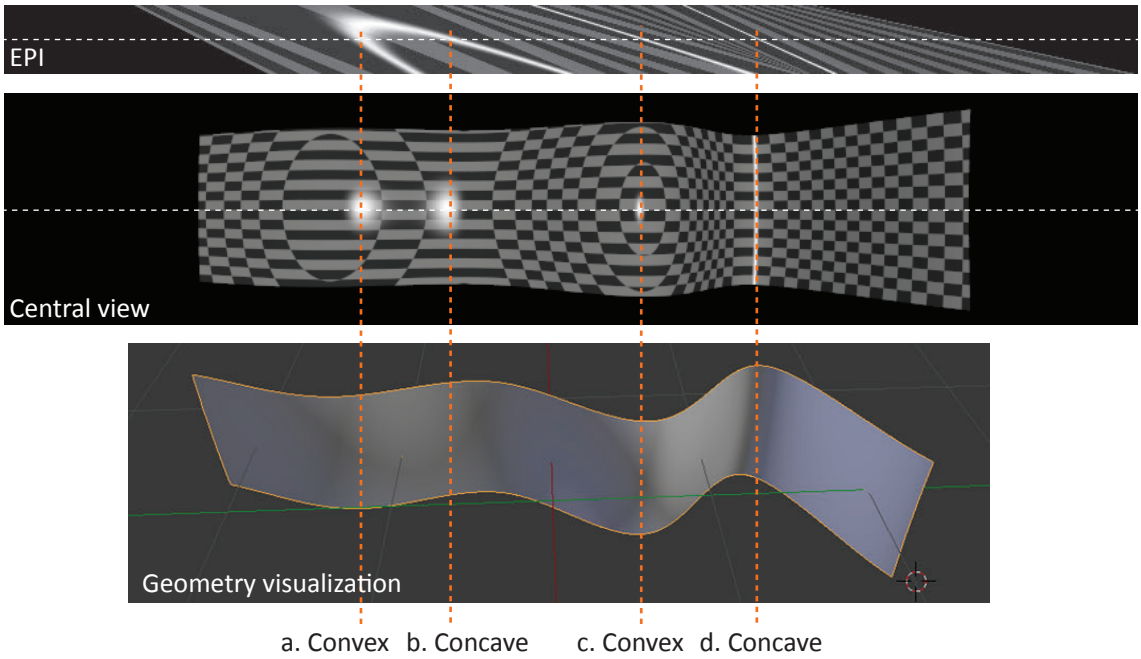


Figure 2.8: Example geometry of specular reflection in EPI. The top image shows an EPI image computed for the central vertical position of the light field. The central view of this light field is given in the middle image. The bottom image shows the geometry of the surface. The white dashed line in the top image corresponds to the white dashed line in the middle image.

features and can have a non linear shape (Figure 2.8, top image). In case of a small baseline the feature track of specular reflection can be considered to be a straight line, where its slope depends on the surface curvature: In case of a convex surface the depth of specular reflection is larger than the one of the object. In case of a concave surface specular reflection appears to be in front of the object [Blake and Bülthoff 1990].

Lin et al. [2002] proposed a method for specular separation and depth estimation from image sequences. First, the mask of specular regions is estimated by evaluating histogram differences between different views. Since the colour of specular regions is different from diffuse counterpart, the depth is estimated at this step reliably only for diffuse points. If this initial estimate gives low reliability of depth estimates for some points, these points are marked as specular as well. In order to find depth values for specular regions a flexible window around the pixel is considered. In the case where the number of non-specular pixels in a window falls below a predefined threshold, the window size is progressively increased. Then, the final consistency measure is computed only over the views where the support window is believed to be mostly diffuse and unoccluded.

Criminisi et al. [2005] suggest to distinguish between specular and diffuse in EPIs by observing that the feature track of a specular reflection for a convex surface will have the slope smaller than the slopes representing diffuse points, but will lie on the top of diffuse points (Figure 2.8a,c). In case of concave surfaces it sometimes not possible to disambiguate the specular and diffuse points (Figure 2.8d).

Stich et al. [2006] formulate a depth map estimation problem for non-Lambertian surfaces observing that an angular sampled image or a feature track in EPI of a specular point varies smoothly. Note, that Chen et al. [2014] rely on the same observation to formulate their bilateral measure.

Tao et al. [2016] estimate specular regions mask and suggest line fitting for these regions instead of computing point colour consistency measure. The assumption is that the colours in SCams are represented by superposition of diffuse colour and colour of a light source, where intensity of specular component varies from view to view. This approach allows to improve depth estimates only on the edges of specular regions. Further, the assumption of a single colour of a light source in the SCam image does not necessary hold. Observe the mixture of light source colours in the SCams marked with dashed lines in Figure 2.9. Typically, the more robust is the approach to the

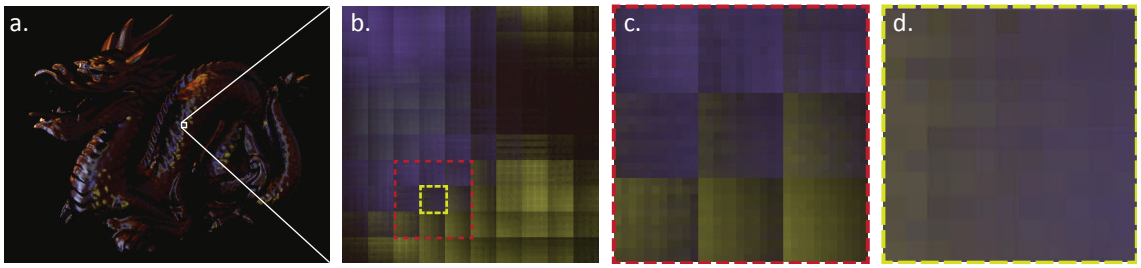


Figure 2.9: Example SCams for a small region of a specular layer of a light field. **(a)** Selected view of a specular layer of a light field. **(b)** SCams computed for each of the points in the regions marked in the view (a). **(c,d)** Close-ups from marked regions in (b).

presence of specular regions the more likely is that the approach will have poorly localised edges with smooth depth transitions. For instance, Heber et al. [Heber et al. 2013, Heber and Pock 2014] model depth from light field as a rank minimization problem and propose a global matching term to measure the warping cost of all other views to the central view. While their method is to a certain extent robust to reflections and specularities, it does not explicitly model occlusions and tends to produce smooth transition on occlusion edges.

2.6.4 Summary

Solving for depth labels on EPIs compared to usage of the information from a full 4D light field is fast. When the global regularization is used, the results obtained by using initial estimates only from EPIs are comparable with one where initial estimates are obtained from the full light field. In cases when depth propagation for textureless regions is performed without global optimization the approaches exploiting the information form the full light field tend to produce more accurate depth estimation results. Finally, robustness of the approach to specular regions is typically achieved as a trade off with precision of edges localization.

2.7 HDR Light Field

Capturing increased dynamic range by either exploiting an advance sensor architecture or multiple sensors combined with a complicated optical system Section 2.3.1 is expensive. To mitigate this problem, the cheaper setups, for example, consisting of several cheaper cameras are also considered in the literature. While the use of multiple cameras allows to obtain a number of views for each moment of time, most of the existing solutions reconstruct only a single view HDR image or video.

The cheapest considered setups consist of two cameras [Lin and Chang 2009, Schmidt and Li 2011]. These approaches might suffer from artefacts in presence of objects close to a camera due to occlusions. The small overlap of dynamic range of the two captured views will also cause artefacts due to difficulties in registration between the views.

The reconfigurable camera setup suggested by Manakov et al. [2013] allows either HDR images or LDR light fields to be recorded, but cannot record an HDR light field.

Another recent work [Keinert et al. 2014] presents the minimal setup for high- dynamic-range video capture. The suggested setup consists of three cameras, where the two outer cameras capture a scene with identical exposure times, longer than the one of the central camera. This allows to extend dynamic range almost twice. The work considers the problem of keeping the setup as cheap as possible. The third camera is required to alleviate the problem of occlusions handling, frequent for two camera setups. The well-exposed pixels from the central view are used directly, the missing dynamic range is reconstructed by warping the side views to the central one. The warping is performed relying on disparity values estimated from the two side views, obtained with the same exposure setting. This method, as well as the one discussed above, allows to reconstruct only a single HDR frame, but not a whole HDR light field.

Capturing HDR-light-field images and videos with spatio-temporal exposure coding was explored by Schedl et al. [2014]. The light-field setup consists of a 7-by-7 camera array. Four neighbouring cameras are performing the recording with 4 different exposure times. Two of the cameras, with the second shortest and the second longest exposure times, are switching their exposure settings between these two values. The algorithm reconstructs four different exposures for each subframe and is designed only for static scenes while the camera motion is allowed. The restriction on static scenes stems from the requirement for point spread functions (PSF) estimation, which are used to remove motion blur in the views captured with longer exposure times. The restriction is also induced by the bandwidth limitations on transmission of subframes: The used camera array cannot capture light-field videos at true video frame rates. The depth is reconstructed using a plane-sweeping algorithm [Zhang and Chen 2004] for each view, taking into account only the views with the same exposure times. Since depth estimation is affected by overexposed and underexposed

pixels in different views, the final depth value is obtained by averaging among four estimates (from four different exposures), taking into account only the estimates with high confidence values. A camera motion path is calculated by matching speeded up robust features (SURF), using iterative closest point (ICP) registration. In order to increase the sampling rate at the longer exposures, the spherical linear interpolation (SLERP) is used. Once all the transformation matrices are computed PSFs are reconstructed. To increase the robustness and to obtain a dense set of PSFs, for each view, the PSFs are clustered by depth values followed by upsampling. Then, within each depth cluster the PSFs are clustered again by their projected shape. The deblurring is then performed using blind deconvolution [Levin et al. 2011], where the estimated PSFs are used as initial guesses. It is noticed that the processed captured views might suffer from ringing and segmentation artefacts, which are though cancelled out during an HDR-frames reconstruction or a final single-image rendering from a light field. Further, due to upsampling and reconstruction of projection matrices during PSFs estimation, it is possible to increase the frame rate of the captured video. The final results might suffer from spatial and temporal artefacts due to interpolation imprecisions.

Further Li and Zhang [2015] obtain HDR-light-field images through multi-exposure technique, where a plenoptic camera is used to capture a light field with a certain exposure value. The authors demonstrate the advantages of the HDR content for the refocusing problem.

The professional LYTRO cinema camera for light-field capturing suggests the 16-bit images representation, which is a good step towards high-quality-HDR light fields. Nevertheless, the cost of this camera is far from being affordable even to semi-professionals.

Chapter 3

Related Work

The work in this thesis is guided by findings in perception of artefacts in temporal domain as well as perception of angularly varying effects. To address these two themes, in Section 3.1 we give a brief introduction into motion perception as well as existing work on trading between ghosting and other image quality factors. Further, in Section 3.2 the work on gloss perception is discussed. Section 3.3 covers material editing problem, where special attention is dedicated to editing in a light field.

3.1 Perception of Ghosting Artefacts

Objects moving through a scene are normally perceived as sharp due to the smooth pursuit of eye motion. This is possibly accompanied by retinal blur of other scene parts that undergo motion relative to the object that is attended to. In video, objectionable ghosting artefacts may be caused by rendering techniques and display technologies such as view blending in image-based rendering [Vangorp et al. 2011, Berger et al. 2010], insufficient frame rate [Watson 2013, Daly et al. 2015], too short exposure times [Stengel et al. 2015], or cross-talk between views in stereoscopic and light field displays [Zwicker et al. 2006, Woods 2012, Wilcox and Stewart 2003].

In each case, ghosting is consistently judged as the key factor degrading image quality [Vangorp et al. 2011, Woods 2012, Wilcox and Stewart 2003]. The reduction of ghosting typically goes hand-in-hand with concessions along other dimensions, for instance a lower image contrast, an increase of motion blur, lower stereoscopic depth range, and so on.

Thus, a previous study by Wilcox and Stewart [2003] assessed the relative importance of brightness and ghosting in a 3D cinema scenario. Here, ghosting is caused by cross-talk between eyes, an issue specific to stereo 3D which depends on object depth and contrast, and is mostly visible on a dark background. Ghosting was simulated by adding various percentages of the left eye signal to the right eye signal. Different brightness levels were created by linearly scaling the original footage down to various target levels. A magnitude estimation task revealed that ghosting was a more critical artefact than reductions in brightness.

In our case, ghosting arises not as a consequence of 3D signal interaction, but as an artefact of HDR reconstruction. Moreover, we are interested (Section 4.1) in relative importance of ghosting to clipping of over-exposed regions and loss of details in under-exposed (while the above study was dealing with overall brightness).

3.2 Perception of Gloss

A number of works have looked into how we perceive gloss; we refer the reader interested in gloss perception to the review by Chadwick and Kentridge [Chadwick and Kentridge 2015]. Here, we are interested in the specific area of how well we can predict the appearance and behavior of glossy surfaces, and how tolerant we are to possible inaccuracies in the depiction of gloss. Reflectance estimation under arbitrary illumination is a highly underconstrained problem [Dror et al. 2001], slightly ameliorated for our visual system if the scene is under natural illumination [Fleming et al. 2003]. This problem is aggravated by the influence of geometry [Vangorp et al. 2007], although it has been shown that specular highlights can aid us in the perception of 3D shape [Fleming et al. 2004]. Further, previous work has shown that our ability to detect inconsistencies in illumination is very poor, tolerating inconsistencies of over 30 degrees in the illumination of different objects in a scene [Ostrovsky et al. 2005, Lopez-Moreno et al. 2010]. Finally, recent works suggest that we do not rely on the physical properties of objects for material perception, i.e., our brains do not perform the inverse rendering pipeline; instead, we rely on invariants or image properties as cues [Fleming 2012, Marlow et al. 2012]. This gives a motivation to our technique for editing in light fields (Chapter 6), where on the one hand the editing should account for the underlying geometry and on the other hand the approximate solution can be sufficient.

3.2.1 Gloss Consistency Between Different Views

One of the key light-field enabled effects are free changes of the view direction, which should trigger a consistent highlight motion over the surface and this way separate the highlight from diffuse texture patterns. In such dynamic conditions the gloss perception is stronger than for static images [Wendt et al. 2010]. When the eye is tracking a highlight the corresponding optical flow vectors become inconsistent with the epipolar line constraint that holds for a diffuse component, which is an important determinant of perceived shininess [Doerschner et al. 2011]. For instance, highlights appear to move behind the diffuse convex surfaces and in front of concave ones [Blake and Bülthoff 1990]. Further, the velocity of a highlight is higher for low-curvature surfaces. While it is an open research problem to what extent any depart from such consistent multi-view highlight behavior can affect the plausibility of gloss effect, it is clear that for suppressed highlight motion the surface may appear more matte and textured rather than shiny [Doerschner et al. 2011]. The agreement in the highlight placement with surface shape increases perceived shininess similar as increasing the highlight size and brightness [Beck and Prazdny 1981]. This indicates that for a consistent roughness increase in a light field (Chapter 6), we need not only to increase the highlight size, but also to propagate highlights between neighbouring views with respect to the underlying geometry. The highlights may become visible in some views with increasing roughness.

3.3 Material Editing

Methods for material editing can be categorized according to what information they rely on. The most advanced techniques rely on having complete scene information, i.e., geometry, lighting, and reflectance properties of objects. In this category, interactive techniques for both parametric [Colbert and Pattanaik 2006, Ngan et al. 2006, Talton et al. 2009], as well as measured [Ben-Artzi et al. 2006] BRDFs were proposed. Some of them address also spatially-varying BRDFs [Lawrence et al. 2006].

For any of such BRDF representations, even sparsely marked user edits must be meaningfully propagated to all image regions of similar appearance [Pellacini and Lawrence 2007, An and Pellacini 2008, Nguyen et al. 2013].

Another group of methods for material editing includes techniques that do not rely on an explicit BRDF representation, but work directly with a 3D scene where the user interactively deforms the shape of reflections or their position on the reflecting surface [Ritschel et al. 2009].

Our work is most related to image-based methods in which the difficulty of humans to perceive inconsistencies in material appearance has been exploited for editing materials [Khan et al. 2006], or even caustics [Gutierrez et al. 2008]. Such approaches benefit from the fact that the scene is observed from a single viewpoint, and thus, some unnoticeable inaccuracies can be present in depth or illumination estimation. These techniques often require additional user interaction [Oh et al. 2001, Yeung et al. 2011]. Meanwhile, other methods assume that per-pixel normal and depth data are given together with the input image, e.g., for shading manipulation [Vergne et al. 2012].

More recently, Boyadzhiev et al. [2015] proposed a purely image-based technique to modify material properties by performing a subband decomposition and selectively sifting its coefficients.

With the increasing interest in light fields, many techniques have been proposed to edit them. These include methods for morphing light fields [Zhang et al. 2002, Wang et al. 2005], interactive deformation of objects [Chen et al. 2005], or propagation of local edits such as painting and scissoring [Seitz and Kutulakos 2002]. Further efforts focused on hole-filling, object reshuffling and resizing, manipulating object depth, and parallax magnification [Zhang et al. 2016]. Recently, some work has targeted automatic edit propagation [Jarabo et al. 2011] (with similar goals as in [An and Pellacini 2008] for BRDFs) to ensure spatial and angular consistency, and different paradigms for efficient light field editing [Jarabo et al. 2014]. Although many new light field editing techniques have been recently proposed, to our knowledge no method exists for gloss appearance editing in light fields.

The main difference between our technique, described in Chapter 6, and the work mentioned above is that our method does not require explicit information about geometry, material, or illumination. Instead, we exploit the fact that such information is encoded in the light field. As a preprocessing step, we compute depth information and a specular-diffuse separation. With respect to the recent work by Boyadzhiev et al. [2015], our technique provides different edits. While they modify high level properties, such as oiliness, glow or blemishes of the skin, we focus on the editing of glossy appearance, and roughness in particular. Furthermore, our approach allows to propagate edits to views of the light field which had no gloss before.

Chapter 4

Motion Aware Exposure Bracketing For HDR Video



Figure 4.1: The HDR video capture algorithm is shown on two consecutive frames (tonemapped for display), separated by a visualization of scene areas that have moved between these frames. These motion areas are analysed on the fly to guide a novel metering algorithm that determines the next exposure, thus minimizing ghosting artefacts. The false colour images on either side show the output of the structural similarity index (SSIM) for each of the RGB channels of the reconstructed HDR frames relative to the ground truth, and indicate that the error in our method is small and well-controlled.

In 2014, sales of digital cameras amounted to 1.8 billion units. Of these, 95% were camera phones. In Q3 of 2014 alone, more than 300 million smartphones were sold (statistics from TomiAhonen Phonebook 2014). This trend continues: Sales of smartphones in Q1 of 2016 amounted to 349 million units, a 3.9 percent increase over the same period in 2015, according to Gartner, Inc. All smartphones and many other camera phones as well as tablets have at least one and often two cameras on board with which images and video can be taken. However, the visual quality obtained with such mobile devices is frequently impacted, as the range of luminance, commonly present in scenes, exceeds that which can be captured. The "HDR mode" on many modern smartphones helps to ameliorate this problem for still images through techniques akin to exposure fusion (Section 2.3.2).

Capturing video in high dynamic range, however, remains a difficult problem (Section 2.3). Mobile devices are particularly ill-suited to directly capture HDR videos due to small sensor sizes that are optimized to capture even higher pixel resolutions. Mobile phones with 4K camera sensors dominate the market over the last few years, as an example, the Sony Xperia series feature 23 MP resolution and the Samsung Galaxy series operate with 12 MP resolution sensors. This means that the capture of strongly dynamic scenes with complex occlusion relations and a high range of luminance values remains a challenge on mobile platforms.

Without resorting to high frame-rate capture, we offer a novel solution that features high reactivity

to changes in the scene content and dynamic range. To this end, we propose a new metering algorithm, which aims to increase captured dynamic range while avoiding the exposures that could lead to ghosting errors during video reconstruction. Advanced metering is feasible for still images (Section 2.3.2), and typically requires a sampling of the full dynamic range of the scene prior to capturing a final small set of exposures. Such an approach is too time consuming to extend to HDR video capture. Thereupon, our metering algorithm does not require the knowledge of the full scene irradiance distribution. In order to obtain an HDR video from a sequence of differently exposed frames we extend the technique [Kalantari et al. 2013] to allow for more freedom in exposure settings, and as such better match it to our metering algorithm

To capture viable HDR video, we therefore present a real-time metering algorithm, and combine it with an adaptive off-line HDR reconstruction algorithm (see Figure 4.1). To guide the design of our algorithm, we have conducted a psychophysical experiment to assess the relative importance of ghosting against dynamic range artefacts. We find that under comparable conditions, ghosting takes precedence over clipped luminance artefacts. Thus, metering is designed such that during capture we avoid ghosting rather at expense of reduced dynamic range, while capturing the full dynamic range when ghosting is unlikely to occur. Through an analysis of motion in the scene, we control the time difference between consecutive exposures, which allows us to minimize motion related artefacts in the resulting videos.

We demonstrate that HDR sequences generated with our system of on-line metering and off-line reconstruction generally feature fewer disturbing artefacts, while at the same time significantly improving the captured dynamic range of the output. The metering part of the algorithm is matched to the capabilities of current mobile devices, and as such, our system would be applicable to current mobile technology and would therefore serve a very large market. In the following sections of this chapter, we describe the perceptual experiment that informed our metering algorithm, the metering algorithm itself and the HDR video reconstruction algorithm. We end this chapter by showing results and drawing conclusions.

4.1 Perceptual Assessment of Ghosting and Over-Exposure Artefacts

Different exposure allocation strategies generally offer a trade-off between different temporal and luminance spacings. If the luminance range covered by the exposures is not large enough, parts of the reconstructed HDR frame may remain over-exposed, while large temporal gaps between exposures are likely to create ghosting in the presence of motion. To evaluate the relative importance of these two artefacts, we perform a psychophysical experiment, based on the subjective assessment methodology for video quality (SAMVIQ) protocol, which is standardized in ITU-R recommendation BT.500 for assessing video quality [ITU-R 2012].

In this protocol, video sequences, including a reference, are shown to the observers. Observers are tasked with assigning a score of visual quality to each sequence with respect to the reference (the continuous range between 1 and 100 is used in our settings.) The reference is additionally shown as one of the ranking sequences. Unlike typical ranking or rating tasks, in SAMVIQ observers can revisit their scores for previously viewed sequences, allowing them to fine-tune their decisions based on later information. This tends to lead to more accurate assessments, reducing noise in the results.

4.1.1 Stimuli

For our experiment, we processed three scenes taken from the Stuttgart HDR database [Froehlich et al. 2014] with the aim of creating new HDR versions of each, with varying amounts of ghosting or over-exposure. LDR frames were extracted from each sequence using different temporal and luminance spacings, which were then re-merged into HDR frames, mimicking standard multi-exposure techniques [Reinhard et al. 2010]. In all cases, three exposures were used per output frame. Either all three exposures were taken from the same frame (0 frames), or the shortest and the longest exposures were taken from the neighbouring frames with increasing spacing (1,2,3 frames). Additionally, four luminance spacings between the extreme exposures were used, set to 3.6, 4.8, 6.0 and 7.2 stops. The sequence created using 7.2 stops of luminance spacing and 0 frames of temporal spacing was selected as the reference. These parameters were selected so that ghosting and over-exposure artefacts affected approximately the same number of pixels at each level, as determined by a pilot study.

The LDR frame is obtained from the ground-truth HDR frame by scaling the latter accordingly with the exposure time. The scaled values are clipped to the range of values representable with an 8 bits per channel image. The extracted exposures were then merged [Debevec and Malik 1997] using linear ramps between 0 and 25 and between 227 and 255 to reduce the weight of unreliable pixels. For each scene, 17 sequences were created.

4.1.2 Experimental Setup

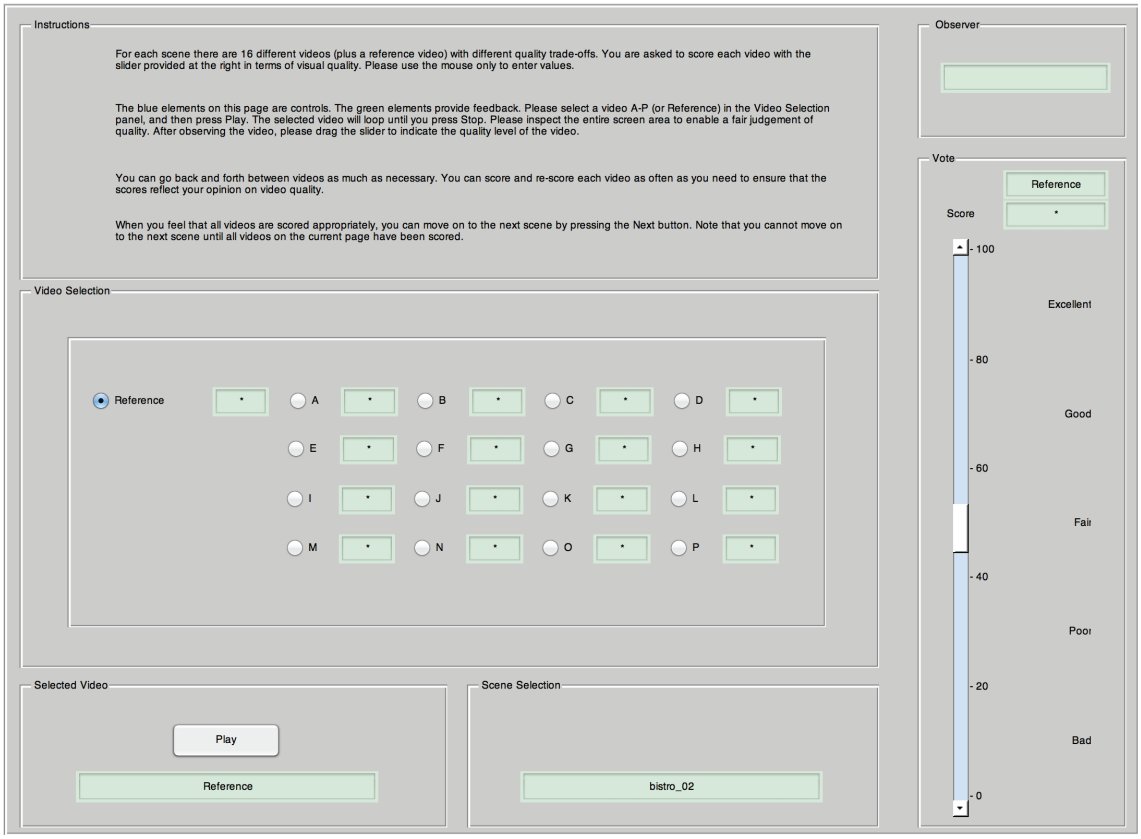


Figure 4.2: User interface used for our psychophysics experiment.

Once all sequences for one scene were scored, following the SAMVIQ protocol, the observers could proceed to the next scene. The order of scenes and sequences was randomized.

The experiment was performed using a dual display setup. Figure 4.2 shows a screenshot of the user interface that implements the SAMVIC protocol [ITU-R 2012]. This interface was shown on a laptop screen (MacBook Pro) and was designed in MATLAB. A secondary HDR display (Sim2 HDR47E-S-4K, HD resolution) was controlled with the aid of the Psychophysics Toolbox [Brainard 1997]. The videos were encoded into the Sim2 LogLuv format, which necessitates encoding into uncompressed format. The MacBook's solid state drives enabled playback of these uncompressed videos at 24 frames per second. To achieve temporal consistency as well as to ensure comfortable display levels, each frame in each sequence was scaled by a constant of 500 to let the peak luminance on the SIM2 display not exceed 500 cd/m^2.

Participants were placed approximately 1m away from the HDR monitor and were allowed to adjust their chair for optimal viewing. A total of 16 participants took part in this experiment (11 male, 5 female), all with normal or corrected to normal vision, with ages between 24 and 60 years ($\mu = 35.1, \sigma = 10.8$).

Nearly all aspects of the SAMVIQ protocol were adhered to. The main differences is that, as shown in Figure 4.2 the interface permanently shows the instructions to the participant.

In total, each observer viewed 51 sequences (4 luminance spacings, 4 temporal spacings, 1 open reference, 3 scenes) and a score was recorded for each of them. Figure 4.3 (a) shows average scores for each of the luminance and temporal spacing combinations across all participants and scenes. Lower scores indicate a lower perceived quality.

4.1.3 Analysis

The obtained results were analysed for significance using ANOVA on each of the main variables (luminance spacing, temporal spacing, scene) as well as on pairwise interactions between them, outlined in Table 4.1. The Pearson correlation of scores against each of the main variables was also computed (last column of Table 4.1). Overall, all three conditions were shown to have a significant effect on the scores, albeit the ghosting artefacts were a more decisive factor in observers' scores, irrespective of the presence of over-exposure.

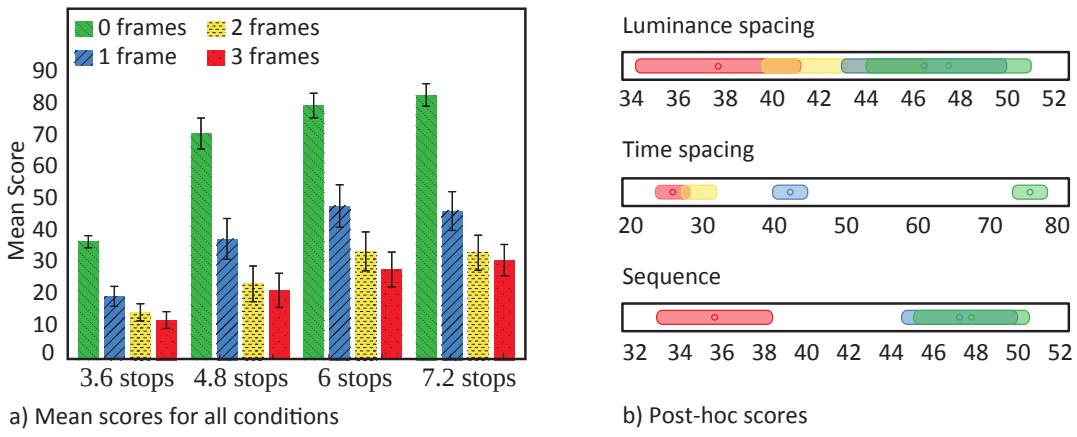


Figure 4.3: Mean scores and errors for each luminance and time spacing condition, averaged over all scenes and participants are shown in **(a)**. The results of post-hoc analysis are shown in **(b)**.

Comparison	F	p-value	Pearson ρ
Temporal	$F(3,764) = 284.5$	$p < 0.001$	-0.67
Luminance	$F(3,764) = 5.2$	$p < 0.005$	0.14
Scene	$F(2,765) = 17.2$	$p < 0.001$	0.18
Temp/Lum	$F(9,752) = 0.2$	$p = 0.989$	
Temp/Scene	$F(6,756) = 5.0$	$p < 0.001$	
Lum/Scene	$F(6,756) = 0.1$	$p = 0.997$	

Table 4.1: ANOVA results for each of the main conditions and interactions between them. The last column shows the Pearson correlation of scores against each of the main variables.

Post-hoc analysis was also performed using the Tukey-Kramer test with 95% confidence interval, shown in Figure 4.3 (b), further supporting the observation that ghosting is deemed a more severe artefact compared to over-exposure.

In HDR imaging, the goal is to extend the dynamic range encoded in the resulting images or videos. However, the above results suggest that this should not be done at the cost of ghosting artefacts, as their presence is likely to nullify the benefit of the extension in dynamic range. Our results complement the previous studies of Masia et al. [2009], where the authors showed that perceived quality of HDR images depends more on the presence or absence of disturbing spatial artefacts than on the exact intensities. In their case the HDR images were obtained from LDR images using reverse tone-mapping operators, and the addressed artefacts are caused by these operators (noise, fringes, etc.). We consider the ghosting artefacts in videos and evaluate the presence of ghosting versus the increased number of details and reduced noise due to the high dynamic range. Further, our study is related to a work [Wilcox and Stewart 2003], discussed in Section 3.1. Though, in our case, ghosting arises not as a consequence of 3D signal interaction, but as an artefact of HDR reconstruction. While we are interested in relative importance of ghosting to clipping of over-exposed regions and loss of details in under-exposed, the study by Wilcox and Stewart [2003] was dealing with overall brightness.

Thus, to maximize the quality of our HDR content, we need to take into account the trade-off between ghosting and over-exposure. As such, to avoid ghosting, we detect and consider motion in the scene while capturing. Our metering algorithm is based on these observations and will be described in the following section.

4.2 Metering Algorithm

Our metering algorithm determines the optimal exposure time for every frame as a function of scene motion and its dynamic range. Based on the psychophysical experiment described in Section 4.1, we select exposures such that ghosting is minimized, while at the same time capturing the highest possible dynamic range.

To achieve that, our algorithm analyses previous frames to determine capture settings for the current frame. In particular, it gathers motion information from the two recent frames to divide the image into regions of local scene motion (local motion regions, or LMRs) and their counterparts, dominant motion regions (DMRs). The separation is performed to ensure that the number of well exposed pixels in DMRs in both frames is sufficient for dominant vector estimation. Its estimation precision influences the appearance of overexposed regions after reconstruction. It is also important to ensure

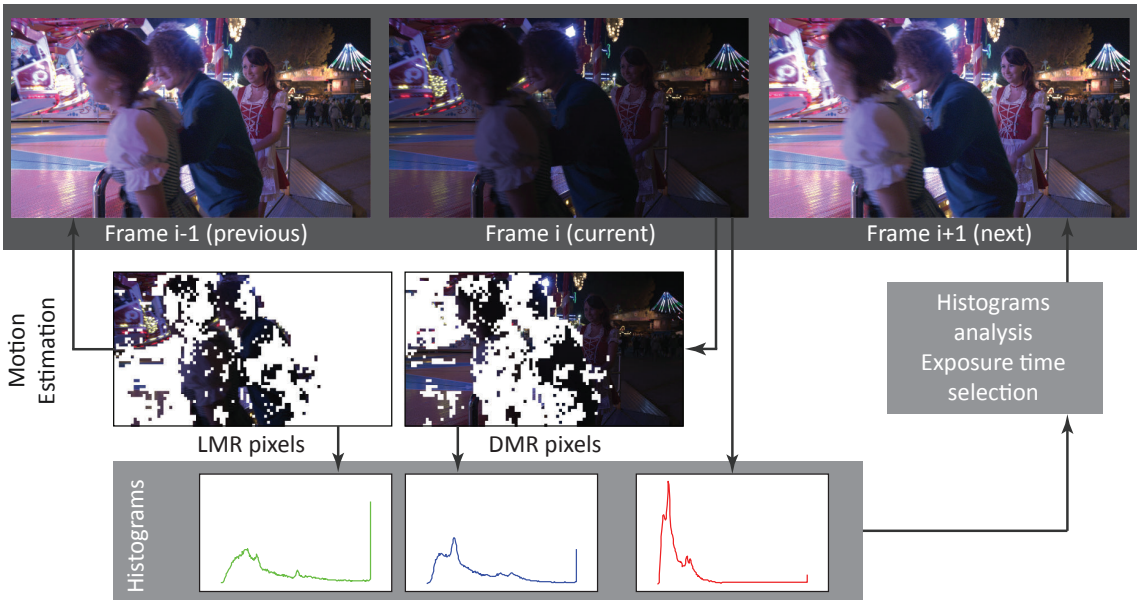


Figure 4.4: A high-level flowchart of our metering algorithm.

that the number of correctly exposed pixels in the LMRs is sufficient for its correct reconstruction.

The normalized histograms of each of these areas, as well as that of the full frame are then computed (normalized meaning that the sum of the values in the histogram bins is one). By examining the extreme bins of these three histograms (full, LMR, DMR) we can assess how well we have captured the dynamic range of the full scene and of the motion regions separately, and therefore decide how the next frame should be exposed. Additionally, as a by-product of the exposure time selection, our algorithm decides in advance whether the exposure for the following frame should be increased or decreased, initializing the next metering cycle. These steps are outlined in Figure 4.4 and will be described in detail in the following sections.

4.2.1 Initialization

To seed the metering algorithm, and to be able to derive the first motion vectors, we initialize the video capture by capturing two frames using the standard metering system provided by the camera or mobile device. In addition we will keep track of whether the next exposure will be longer or shorter, allowing us to successively step up and down through the luminance range. We set the initial direction to 'shorter' as over-exposed regions carry no useful information for correspondence computation in the HDR video reconstruction phase.

4.2.2 Motion Estimation

To estimate the motion between frames, we use the hierarchical diamond search (HDS) algorithm [Urban et al. 2009], a real-time block matching algorithm, which performs a recursive search over a multi-level image decomposition with a diamond shape window around the current position. The HDS algorithm returns one motion vector for each block. Since subsequent frames are differently exposed, we adjust pixel intensities in the frame with the shorter exposure to the longer exposed frame.

As a distance measure between blocks we use an absolute difference between mean gamma corrected pixel intensities of the blocks, rather than the standard sum of absolute differences, as it is faster to compute and better in compensating for the differences between frame exposures.

The HDS algorithm is a combination of two approaches: enhanced predictive zonal search (EPZS) [Tourapis 2002] and hierarchical motion estimator (HME) [Chupeau 1993]. It combines the fast performance of EPZS, achieved due to recursive diamond search window, with the precision of HME, which allows to find large motion vectors due to hierarchical image decomposition. In other words, the HDS algorithm performs a recursive search over a multi-level image decomposition with a diamond shape window around the current position.

For each block HME relies on the predictors obtained from the neighbouring blocks and previous frames. In HDS multi-level image decomposition provides robust hierarchical predictors. These predictors ensure estimation of large motion vectors. The spatial predictors provide the homogeneity of the motion field.

The refinement step recursively selects small displacements around the current best motion vector. This local search is initialized with the results of the predictive step. The best position is chosen as the new search center of the recursive process. If the best position is already at the center of the pattern, the search stops.

The dominant motion vector for a frame is estimated with a block-based recursive weighted least-squares method after the motion field is computed [Urban et al. 2009]. An affine model is used in which the dominant motion vector consists of 6 components (translation (2) + rotation (4)). Homogeneous blocks (particularly overexposed areas) are excluded from the computation of dominant motion vectors. A block is marked as homogeneous if the vertical or the horizontal gradient is below a set threshold (80 in our case). Blocks which do not follow the dominant motion are marked as LMRs.

4.3 Histogram Computation

We observe that ghosting artefacts are typically the consequence of local motion, particularly in high contrast regions, which pose difficulties for registration and HDR reconstruction algorithms. Moreover, the stronger the motion the higher is the probability of ghosting. Motion estimation algorithms likely find the correct motion vectors for regions with little motion between subsequent frames, while for regions with large motion vectors a sufficient amount of well-exposed pixels in subsequent frames should be ensured. As such, when computing the histogram of LMRs, each pixel is weighted according to its motion magnitude. The magnitude of local motion is computed after subtracting the dominant motion vector from the full motion field.

Additionally, when the camera is static (zero dominant motion vector), the observer is sensitive even to small local motions. In this case, instead of direct pixel weighting with the local motion magnitude, we employ a non-linear concave mapping of such magnitudes as the weights (Figure 4.5). This increases the weights of small motions, but keeps the priority for larger motions.

In the presence of camera movement, motions that are relatively small in comparison with the dominant motion are not easily distinguishable, and thus we apply a convex mapping (Figure 4.5). This effectively suppresses the weights of the regions with small motions and allows to give the priority to dynamic range capturing.

We do not consider the motions which have motion amplitudes larger than 80 pixels. To compute

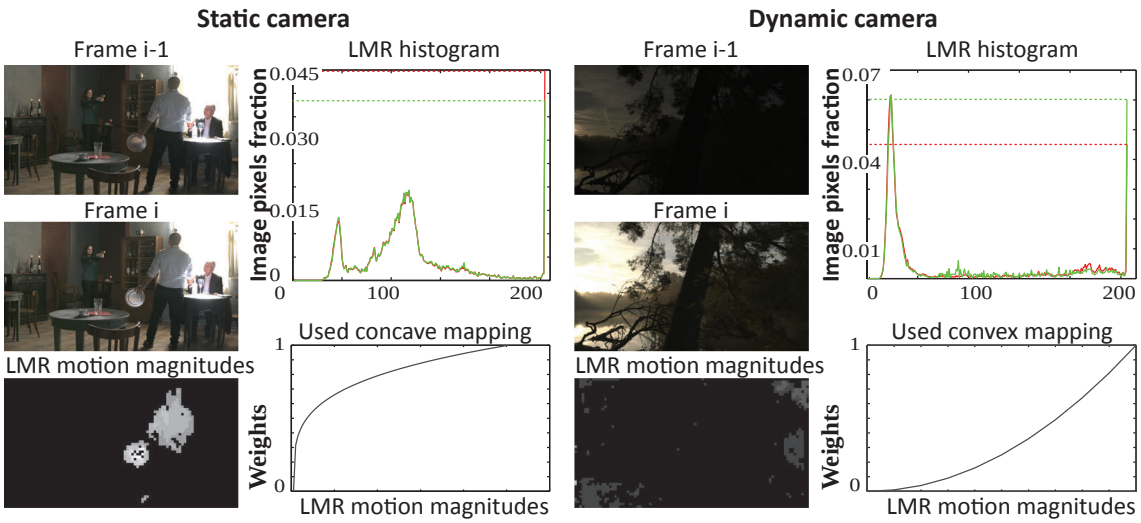


Figure 4.5: An example of the LMRs histogram computation. The red curve corresponds to the case of the histogram computation with accounting each pixel in LMRs with weight one. The green curve corresponds to the histogram obtained by weighting each pixel in LMRs with weight computed from its motion magnitude. The dashed lines show the percentage of overexposed pixels in each case. The example on the left shows an example of static camera. The example on the right shows an example of moving camera.

weights we normalize the motion amplitudes to the range [0,1], dividing by 80. In case of concave mapping we use the mapping $w = m^2$ and for convex we use $w = m^{1/4}$, where w is the weight for the pixel with motion magnitude m. The absence or presence of dominant motion between frames is a binary decision based on the thresholds for each of the dominant motion vector components.

4.3.1 Exposure Time Selection

The goal of this step is to find an exposure time t_{i+1} for the next frame $i+1$, ensuring that the following two conditions are fulfilled:

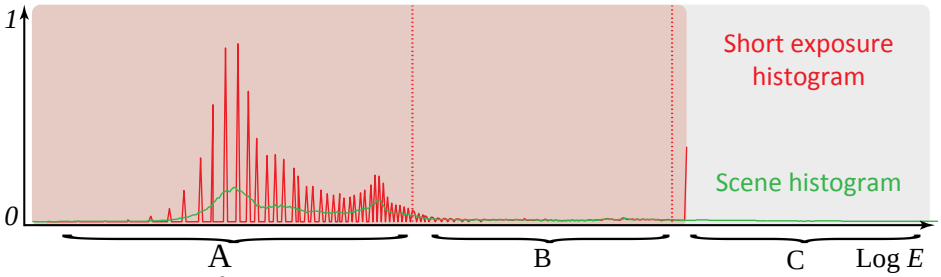
C1. *If the pixels are over-exposed (under-exposed) in the current frame and we decrease (increase) exposure time, they should not be under-exposed (over-exposed) in the next frame. This means that these pixels will be properly captured at least in one of two subsequent frames.*

C2. *The next frame should contain enough meaningful information in DMRs and LMRs for dominant and local motion estimation. This means that a certain percentage of pixels should be properly exposed in both frames.*

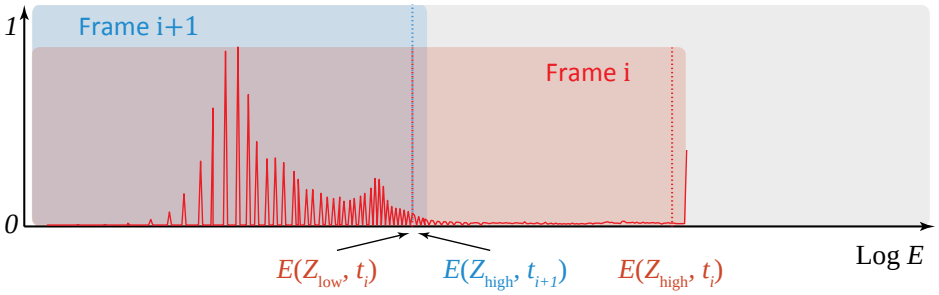
Effectively, this means that the dynamic range captured by two subsequent frames is extended as much as possible, while ensuring that sufficient correspondences will be found by the optical flow algorithm in the HDR reconstruction phase. To that end, we first calculate the largest exposure time increment (or decrement) possible, and then refine this initial estimate based on additional criteria.

Initial Exposure Time Estimate. Exposures with increasing exposure times capture increasing ranges of scene irradiance, albeit that all start at 0 W/m^2. Nonetheless, a good portion of this data is lost due to under-exposure, noise and over-exposure (Figure 4.6 (a)). We therefore define

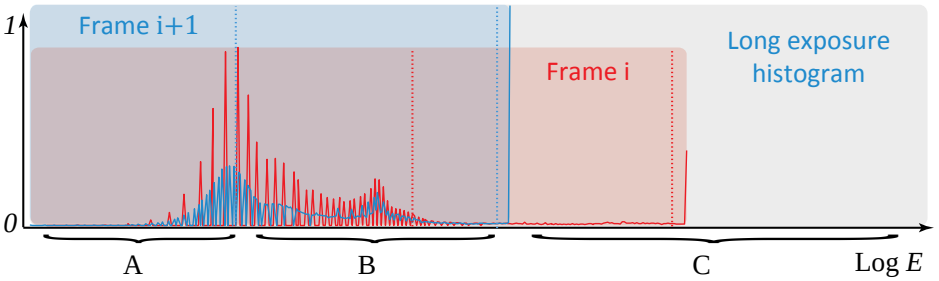
two thresholds $Z_{\text{low}} = 0.2 \cdot 255 = 51$ and $Z_{\text{high}} = 0.9 \cdot 255 \approx 230$, which approximate the lower and upper pixel value boundaries of the well-exposed region of the current frame's histogram (Figure 4.6 (a and b)). These boundaries are also matched to similar values used in the HDR video reconstruction algorithm. Pixels with values higher than Z_{high} will be reconstructed from frames with shorter exposure times, and pixels with values lower than Z_{low} will be obtained from frames with longer exposure times.



a) Capture range of a single exposure: A - Unreliable due to sparse sampling. B - Reliably captured. C - (Near-) Over-exposed, collected in last bins.



b) Matching of $E(Z_{\text{low}}, t_i)$ to $E(Z_{\text{high}}, t_{i+1})$.



c) Capture range of the next frame with the time obtained after exposure time adjustment. Note that in dark regions the discretization error is reduced and scene histogram is approximated much better due to dense sampling of that area.

Figure 4.6: Rationale for our initial exposure selection strategy **(a - b)**, and the result of the exposure adjustment **(c)**.

Here, we describe the search process to establish an exposure time t_{i+1} that will be longer than exposure time t_i. The search for a shorter exposure time is analogous.

The estimate of irradiance E, corresponding to pixel value Z in the frame with exposure time t, can be calculated as:

$$E(Z,t) = f^{-1}(Z)/t. \tag{4.1}$$

where $f^{-1}()$ is the inverse camera response curve (which can be computed using standard techniques

[Debevec and Malik 1997]). Then, the following is required to fulfil condition C1:

$$E(Z_{\text{high}}, t_{i+1}) \geq E(Z_{\text{low}}, t_i) \tag{4.2}$$

From Equations (4.1) and (4.2) we calculate an upper bound $\bar{t}_{i+1}$ on the exposure time of the next frame:

$$t_i < t_{i+1} \leq \bar{t}_{i+1} = f^{-1}(Z_{\text{high}}) t_i / f^{-1}(Z_{\text{low}}). \tag{4.3}$$

We initialize the estimate $\hat{t}_{i+1}$ of the optimal exposure time t_{i+1} with its upper bound: $\hat{t}_{i+1} = \bar{t}_{i+1}$.

Exposure Time Refinements. To prevent ghosting in subsequent reconstruction, we may have to adjust the current estimate $\hat{t}_{i+1}$ (Figure 4.6 (c)). Here we describe how we ensure that the condition C2 is fulfilled.

From the histogram of the current frame we estimate $N_{\text{LMR,over}}(\hat{t}_{i+1})$ and $N_{\text{DMR,over}}(\hat{t}_{i+1})$ which are the expected percentages of over-exposed pixels in LMRs and DMRs of the next frame, as function of the new exposure time $\hat{t}_{i+1}$. We require these values to be below the predefined thresholds $\rho_{\text{LMR,over}}$ and $\rho_{\text{DMR,over}}$ (which are computed as described in Section 4.3.3):

$$\begin{aligned} N_{\text{LMR,over}}(\hat{t}_{i+1}) &\leq \rho_{\text{LMR,over}}, \\ N_{\text{DMR,over}}(\hat{t}_{i+1}) &\leq \rho_{\text{DMR,over}}. \end{aligned} \tag{4.4}$$

The estimate $\hat{t}_{i+1}$ can take values between $\bar{t}_{i+1}$ and t_i. As such, starting at $\hat{t}_{i+1}$, the exposure time is estimated by testing all possible values within this range until Eq. 4.4 is fulfilled. If t_i is reached and Eq. 4.4 is not fulfilled, it means that the step in a given direction can not be performed.

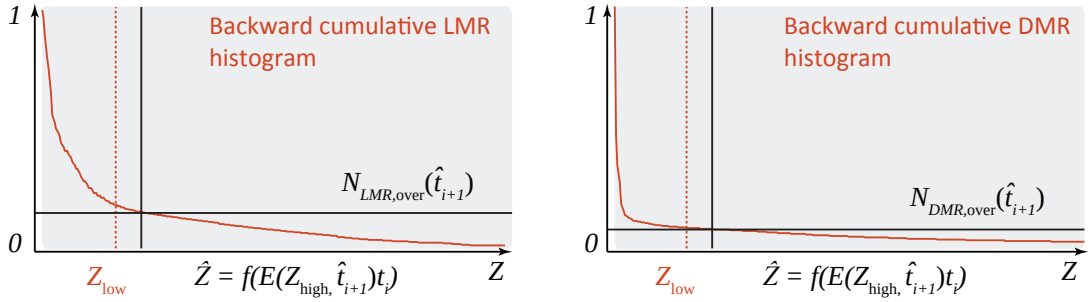


Figure 4.7: Backward cumulative histograms of LMRs and DMRs. The number of over-exposed pixels is equal to the histogram value at the bin with index $\hat{Z}$.

The expected number of over-exposed pixels in the next frame equals the number of pixels with irradiance values higher than $E(Z_{\text{high}}, \hat{t}_{i+1})$. This corresponds in the current frame to the number of pixels with pixel values higher than $\hat{Z} = f(E(Z_{\text{high}}, \hat{t}_{i+1})t_i)$. Given the histograms of LMRs and DMRs, and the value of $\hat{Z}$, we can immediately get $N_{\text{LMR,over}}(\hat{t}_{i+1})$ and $N_{\text{DMR,over}}(\hat{t}_{i+1})$ by computing the backward cumulative histograms h_{LMR} and h_{DMR} (or standard cumulative histograms in the case of exposure decrease) with the number of bins equal to the frame bit depth (Figure 4.7). The number of over-exposed pixels is then equal to the histogram value at the bin with index $\hat{Z}$: $N_{\text{LMR,over}}(\hat{t}_{i+1}) = h_{\text{LMR}}(\hat{Z})$ and $N_{\text{DMR,over}}(\hat{t}_{i+1}) = h_{\text{DMR}}(\hat{Z})$.

In many non-professional (particularly mobile) cameras, exposure settings can only be adjusted using a given exposure compensation step s. Based on that, the number N of the possible settings of $\hat{t}_{i+1}$ is predefined for a particular camera. The value of N can be found from $\overline{V}_{i+1} = V_i + Ns$,

where $V_i = \log_2(t_i/t_0)$ is an exposure value, and t_0 is the time of the frame with zero exposure value (optimal exposure time for a scene). Then from Eq. 4.3 we derive:

$$\log_2\left(\frac{t_i}{t_0}\frac{f^{-1}(Z_{\text{high}})}{f^{-1}(Z_{\text{low}})}\right) = \log_2\left(\frac{t_i}{t_0}\right) + Ns. \tag{4.5}$$

Note, that $\log_2(t_i/t_0)$ on the left and right hand side cancel out, and N depends only on the camera response curve, exposure compensation step and the values of Z_{low} and Z_{high}. The bin numbers $\hat{Z}_j$, $j \in [0,N]$ that correspond to each of these settings can be precomputed:

$$\hat{Z}_j = Z_{\text{low}} 2^{(N-j)s}. \tag{4.6}$$

This simplifies the adjustment step to cumulative histogram computation, and in the worst case requires $2N$ comparisons.

4.3.2 Exposure Time Direction

To compute the exposure time for the next frame we first need to determine if we want to decrease or increase exposure time. This can be done by analysing the data from the current frame or it can be derived from the exposure selection step for frame $i-1$. Here we use the second approach, as it does not require any additional analysis.

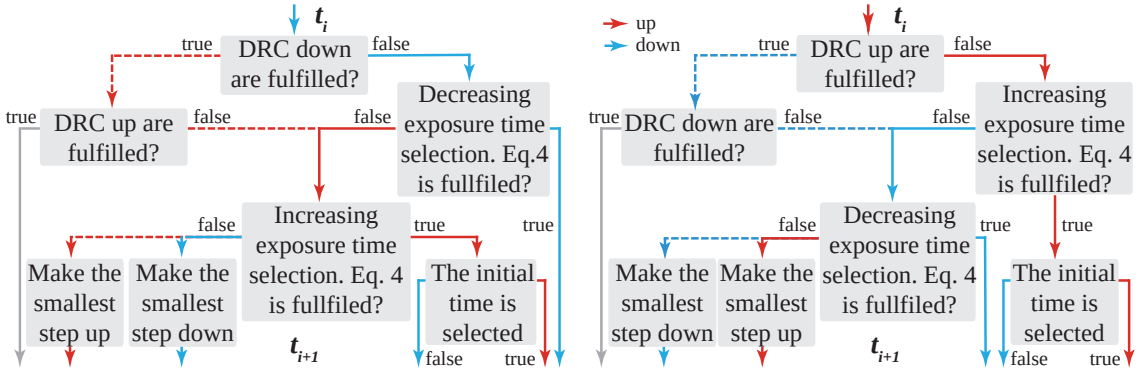


Figure 4.8: The decision tree for selection of the next exposure time dependent on the initial direction: decreasing (left) or increasing (right). A by-product of exposure time selection for the next frame is the initial exposure direction for the frame with index $i+2$, and is given by the colour of the output arrow. The grey arrows denote no changes in the exposure time and direction. Dashed lines denote the branch which corresponds to the case when the DRC in the initial direction is fulfilled.

Before executing the exposure time selection step for frame i as described previously, we check whether the full scene irradiance range is already captured in the specified direction by evaluating the last or first bin of the full frame histogram. Additionally, we check whether the camera settings allow us to make a step in that direction. We refer to these two conditions as *dynamic range conditions (DRC)* (Figure 4.8).

In the simplest case, if the scene's full dynamic range is already captured in both directions we do not change the exposure time and keep the initial direction. This would occur for scenes that can be captured with a single exposure.

Either the DRC or failure to fulfill the conditions in Eq. 4.4 may prevent an exposure time change in the current direction. In that case, the direction is reversed. If any condition in Eq. 4.4 prevents

exposure adjustment in both directions, then the exposure time is changed in the current direction by the smallest step allowed by the camera. This helps to expand the captured dynamic range. Specifically, the smallest step in the initial direction is chosen if the DRC in that direction is not fulfilled, and in the opposite direction, otherwise.

Finally, if the current direction indicates an exposure time increase, we check if an adjustment from the initial time estimation is required, as explained in the preceding section. If that is the case, then for the next frame, the direction will be reversed to exposure decrease to prevent too large overexposed areas in LMRs with no data. For an exposure time decrease this step can be omitted. This asymmetry in operation is due to the fact that any exposure length will result in capture of information in the darks (albeit possibly with noise), whereas bright regions may get clipped, dependent on exposure duration. The full decision process for the direction of the next exposure change is illustrated in Figure 4.8.

4.3.3 Thresholds

We formulate condition C2 (Section 4.3.1) in terms of the threshold percentages ρ_{DMR} and ρ_{LMR} of under- and over-exposed pixels in LMRs and DMRs. Additionally, the thresholds for LMRs depend on the percentage A of pixels in LMRs with respect to all pixels in the frame. The rationale behind such threshold choice is that the exposure selection should not be optimized for tiny moving regions, where ghosting artefacts would be less disturbing than large clipped image areas. Note that the visibility of ghosting typically depends on the contrast of moving objects with respect to the background [Daly et al. 2015, Stengel et al. 2015]. This is implicitly covered by our algorithm. Whenever a dark (bright) object moves over a bright (dark) background, both the background and the object are included into the LMR. Since our algorithm gives higher priority for over-exposed regions, the ghosting artefacts in these regions will be minimized.

$\rho_{LMR,over}^{A_{high}}$	5%	$\rho_{LMR,over}^{A_{low}}$	90%	$\rho_{DMR,over}$	20%
$\rho_{LMR,under}^{A_{high}}$	85%	$\rho_{LMR,under}^{A_{low}}$	100%	$\rho_{DMR,under}$	90%

Table 4.2: The threshold percentage ρ_{LMR} and ρ_{DMR} of under- and over-exposed pixels in LMRs and DMRs. Note that for LMRs different thresholds are used as a function of relative size A of these regions with respect to the whole frame.

The threshold values used in our implementation are summarized in Table 4.2 and were chosen to minimize ghosting. This is consistent with the findings of our experiment in Section 4.1. Here, we compute the size A of the LMRs relative to the image size, and express this as a percentage. If this percentage is smaller than $A_{low} = 0.01\%$, then the LMRs are considered insignificant and therefore the algorithm will give priority to maximize the capture of the scene's dynamic range. Similarly, this percentage is greater than $A_{high} = 1\%$, then ghosting is likely to be noticeable and therefore priority is given to avoid it. Effectively, a higher priority of motion leads to a lower percentage of allowed under- and over-exposed pixels in LMRs. For LMRs with a relative area inbetween A_{low} and A_{high} the thresholds are determined through linear interpolation of the threshold values in Table 4.2:

$$\rho_{LMR} = (\rho_{LMR}^{A_{high}} - \rho_{LMR}^{A_{low}}) \frac{A - A_{low}}{A_{high} - A_{low}} + \rho_{LMR}^{A_{low}}. \tag{4.7}$$

4.3.4 Discussion

Our algorithm could either adjust the exposure time or manipulate the ISO value, which could be used, for instance, to reduce blur [Narasimha and Batur 2012]. In that case the ISO and exposure time are adjusted based on the amount of motion. Thus, when strong motions happen in dark areas of a scene, instead of increasing exposure time the ISO value could be increased. Note that if motion occurs in mid to high luminance areas our algorithm will refrain from selecting long exposure times to avoid overexposure of moving objects.

4.3.5 Implementation

The algorithm was implemented in C++. To achieve real-time performance, the typically HD input resolution is downsampled to 40%. In the current implementation, motion estimation takes from 12–25 ms, exposure adjustment 2–3 ms, histogram computation 1 ms, and exposure time estimation less than 1 ms. All measurements have been performed on an Intel Xeon CPU E5-1607 processor.

Our algorithm incorporates a real-time block-based motion estimation solution as well as efficient histogram analysis. It is therefore well suited for mobile capture applications where resources are limited. Android 5 (Lollipop) allows capture settings to be adjusted per frame, whereas iOS 8 exposes manual camera controls through its API, which allows exposure times to be set for each frame, with probable delay of several frames from the moment the request about exposure time change is sent. Since the content of the scene typically changes smoothly, this delay should not have the strong affect on the performance. Our algorithm can readily be ported to these environments.

4.4 HDR Video Reconstruction

Our real-time metering algorithm allows the production of a sequence of exposures with exposure times that step up and down by some number of frames determined by an analysis of each frame. Moreover, the exposures do not have a regular spacing in exposure time, even if they are captured with a constant frame-rate. Current HDR video reconstruction algorithms on the other hand expect an evenly spaced sequence of exposures that only step in one direction every N frames. As such, they cannot cope with our configuration of exposures, and therefore there is a need to develop a new algorithm that is matched to this input.

We find, however, that we can repurpose an existing state-of-the-art algorithm, which has shown to perform well on evenly spaced exposure sequences. In particular, we adapt Kalantari's algorithm [2013], which has the strong advantage that it is able to reconstruct an HDR video without loss of frame-rate relative to the input exposure sequence.

Algorithm. We refer to an input LDR frame as the *reference frame*. In its simplest form, one could view HDR reconstruction as an iterative two-step process. First, for a given time-step, we synthesize a set of *exposures* such that the reference frame becomes one exposure of an exposure stack. In the second step the exposure stack can then be merged into an HDR frame, using the previous estimate of the HDR frame to synthesize better exposures, which are merged into an improved HDR frame. This process is iteratively refined.

During exposures synthesis, firstly, all the *missing exposures* (the exposures for the frame, excluding

the reference frame itself) are extracted from the previously reconstructed HDR frame. In the very first iteration, the missing exposures are created by adjusting the exposure of the reference frame. Then, the missing exposures are refined from the exposures in the exposures stacks of adjacent frames, using a search and vote procedure ([Kalantari et al. 2013, Sen et al. 2012, Simakov et al. 2008]). The exposures used for the refinement are called *source frames*.

Exposure Arrangement. In Kalantari et al. [2013] reference frames are captured using a repeating sequence of monotonically increasing exposure times. This means that every N frames, the exposure time resets, causing a large jump in exposure time. For those frames, optical flow may be computed less reliably. In contrast, we allow exposure times to oscillate, therefore ensuring that two consecutive reference frames will always have a sufficient number of overlapping pixel values to compute correct optical flow.

Exposure Times Stack Selection. Before the iterative process, the exposure times for the exposure stack of each frame should be determined. The exposure times of frames within the sequence to which the frame belongs to and two neighbouring sequences are immediately added to the exposure times stack. Additional exposure times are obtained by *recursively searching* separately for the shorter and longer exposure times of reference frames to the right and to the left of the current frame (direction to the right denotes the direction to the first reference frame and direction to the left corresponds to the direction to the last reference frame). The found exposure times are merged into a single stack. The source frames for each exposure are one or both adjacent frames, to the left or to the right. The directions to the source frames coincide with the directions to the reference frame at given exposure level.

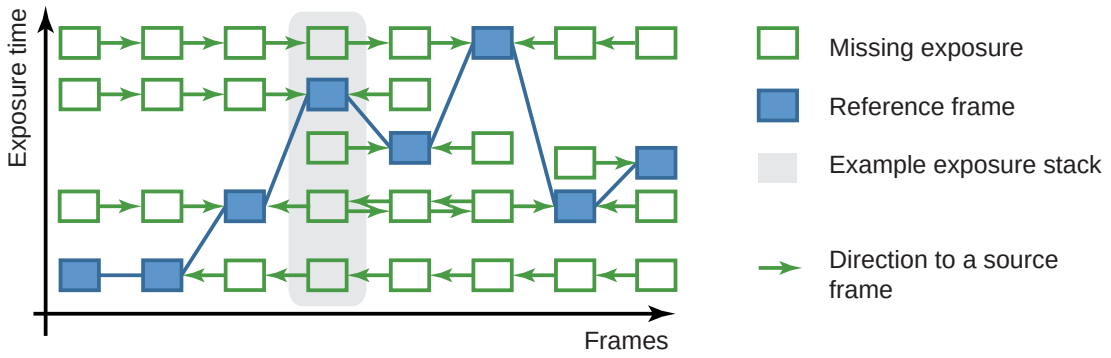


Figure 4.9: Example reference frames arrangement with their exposure stacks.

Recursive Search. Here we describe how the recursive search for the shorter exposure times from frames to the right of the current frame i is performed. Around the current frame there exists a range of frames $K = [i, r]$ that have sufficient spatial overlap with the current frame. The search for the frame f^* with the shortest exposure time in this range, which is shorter than the exposure time of i, is performed. We store the found exposure time and set $r \leftarrow f^*$. The process is recursively repeated until either K is empty or found exposure time has sufficient overlap with exposure time of i (as defined in Section 4.3.1 C1). This process results in the exposure arrangement similar to Figure 4.9.

In the base version of our algorithm the range K is limited to the sequence to which the frame belongs to and two neighbouring sequences. This strategy is sufficient for most scenes, but might cause large bright areas to over-expose if there is motion in dark parts of the scene concurrently.

To obtain the temporal coherence with the base version all the frames could be clipped to the dynamic range of the frame with the smallest dynamic range.

Merging Process. In our exposure arrangement each exposure could have only one source frame, the right or the left one. At the same time, the same pixel could be captured properly in several exposures, and reconstructed better or worse from different source frames, as a result of the search and vote process. To account for this we modify the exposure merging scheme of Debevec and Malik [1997].

After the search and vote process, for the current frame i, we have a set of complementary exposures $I_{n,i}$ with exposure times t_n, $n \in [1,N]$. At the first iteration, for each pixel p_n in each exposure $I_{n,i}$ we calculate a weight $\alpha_{p,n}$ that represents its level of unreliability. Setting these weights follows Kalantari's scheme [2013], with $\alpha_{p,n} = 1$ for well-exposed pixels, a linearly decreasing ramp for pixel values above Z_{high}, and a non-linearly increasing ramp for pixels below Z_{low}. The non-linearity is adjusted based on the amount of motion of pixel p_n. Additionally, we have available the probability of a correct match R_n that stems from the patch-based search and varies at each iteration.

As discussed above, some exposures may have the same exposure time, with contributions coming from previous and next frames. We replace such exposures $I_{n_1,i}$ and $I_{n_2,i}$ on the exposure stack with a new exposure $I_{n',i}$ which is a weighted average of both input exposures:

$$I_{n',i} = \frac{\sum_{n\subset[n_1,n_2]}(1-\alpha_{p,n})R_n I_{n,i}}{\sum_{n\subset[n_1,n_2]}(1-\alpha_{p,n})R_n}. \tag{4.8}$$

The new weight for this exposure is $\alpha_{p,n'} = \min(\alpha_{p,n_1}, \alpha_{p,n_2})$. The resulting N' exposures are merged with a standard technique [Debevec and Malik 1997], albeit that we add our pixel reliability measure into the weighting scheme:

$$I_{\text{hdr},i} = \frac{\sum_{n\in N'}(1-\alpha_{p,n})\,\Lambda(I_{n,i})\,I_{n,i}/t_n}{\sum_{n\in N'}(1-\alpha_{p,n})\,\Lambda(I_{n,i})}. \tag{4.9}$$

Here, $\Lambda()$ is a hat function that reduces the contribution for high and low pixel values. For frame i this constitutes the HDR output of each iteration, and after sufficient iterations have been performed, the final HDR output.

4.5 Results and Limitations

For evaluation we apply our metering algorithm to HDR sequences from a publicly available database [Froehlich et al. 2014] and compare the reconstructed result with ground-truth HDR video. The exposures are extracted from the HDR frames as discussed in Section 4.1.1, where additionally gamma correction is applied before clipping. To compare the results, we used the structural similarity (SSIM) index [Wang et al. 2004], applied between any of the reconstructed HDR videos and the input HDR video. SSIM was chosen to assess structural differences between frames.

Before comparing against the current state-of-the-art, we first show the effect of some of our design decisions. Our merging approach has the option of adding extra exposures to increase temporal stability (as discussed in Section 4.4: Recursive Search). We computed the dynamic range for each frame in a reconstructed sequence as $\log(L_{\max}) - \log(L_{\min})$. This was done with the base and full versions of our algorithm for two sequences. The dynamic range in both cases is 11 stops, but the variance is more than 7.4 times smaller with the full version (0.126 against 0.017).

We further compare our approach with Kalantari's [2013] and following their evaluation, we experimented with two exposure values at -3 and 0 stops and with three exposures at -2, 0 and 2 stops. Here, zero indicates the optimal exposure value for each scene.

We demonstrate our results by comparing close-ups from sequences taken from an HDR video database [Froehlich et al. 2014] (Figures 4.13 — 4.10), as well as the Tunnel sequence (Figure 4.15). The HDR frames obtained with our and Kalantari et al. [2013] approaches are tonemapped for display with Mantiuk's operator [2008], except for the Fireplace sequence in Figure 4.11, which was tonemapped with a sigmoidal operator [Reinhard and Devlin 2005]. For selected HDR frames we show the structural similarity index (SSIM) [Wang et al. 2004]. Thus, in each of these figures, from left to right we show:

1. A frame reconstructed with Kalantari's method, with 2 alternating frames at -3 and 0 f-stops. Next to it are two SSIM comparisons of Kalantari's result with the HDR ground truth, the top image shows the structural component of the SSIM metric. The bottom image shows the full SSIM metric.

2. Another frame reconstructed with Kalantari's method, this time with a repeating pattern of three exposures, set at -2, 0 and 2 f-stops, followed by the SSIM results for this frame.

3. The third frame is reconstructed with our full method, taking motion into account. The two panels next to it show the SSIM differences with respect to the ground truth, as before.

4. The last image in each row shows a tonemapped version of the ground truth.

In all cases, our algorithm produces a closer match to the ground-truth. In most cases the differences with the ground-truth are structural in nature.

Comparing the structural component of SSIM with the full SSIM (which additionally includes an assessment of luminance and contrast differences), we see that for most results in this document these two results are subjectively similar. This indicates that the main reduction of visual quality is due to changes in the structural component. Clipping affects both contrast and luminance, and therefore the SSIM differences are only in part due to structural differences. Extreme ghosting may also affect contrast and luminance, as seen in Kalantari's results in the bottom row of Figure 4.13. In dark parts of a scene with significant motion in the bright parts, the SSIM metric reveals that degradation can be almost entirely due to contrast and luminance issues. See for instance, the bottom row of Figure 4.11. Finally, Figure 4.14 shows that in scenes with motion blur, we adequately reproduce the HDR ground-truth, without introducing artefacts.

To quantify these results, we calculated mean SSIM values for 334 frames drawn from 7 sequences for each method (Kalantari 2 and 3 exposures, and ours) and 2 SSIM scores (full metric, structural component only). These scores were subjected to a one-way ANOVA, showing highly significant differences between the three algorithms ($F(2,1001) = 25.54, p < 0.0001$ for the full SSIM metric and $F(2,1001) = 22.44, p < 0.0001$ for the structural part). The lower mean scores in Figure 4.16

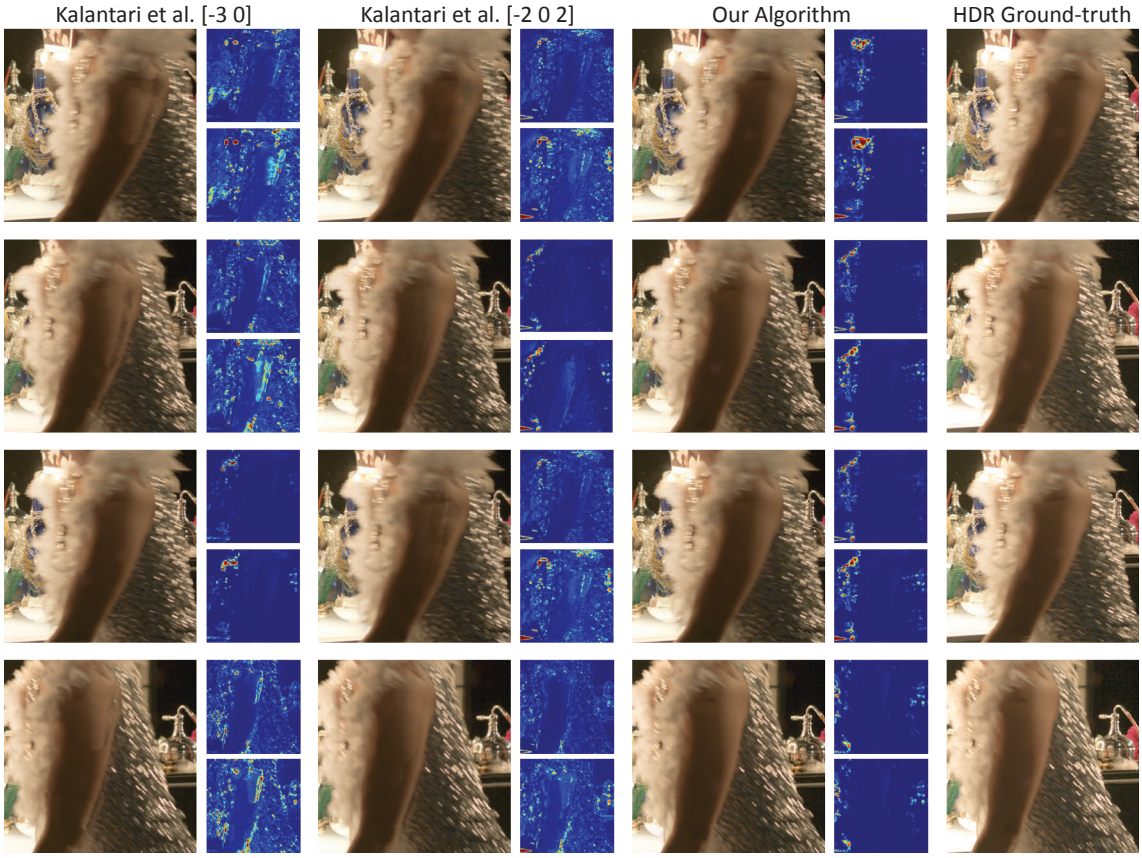


Figure 4.10: Comparisons for four different frames from the Showgirl sequence. See main text for an explanation of the various panels.

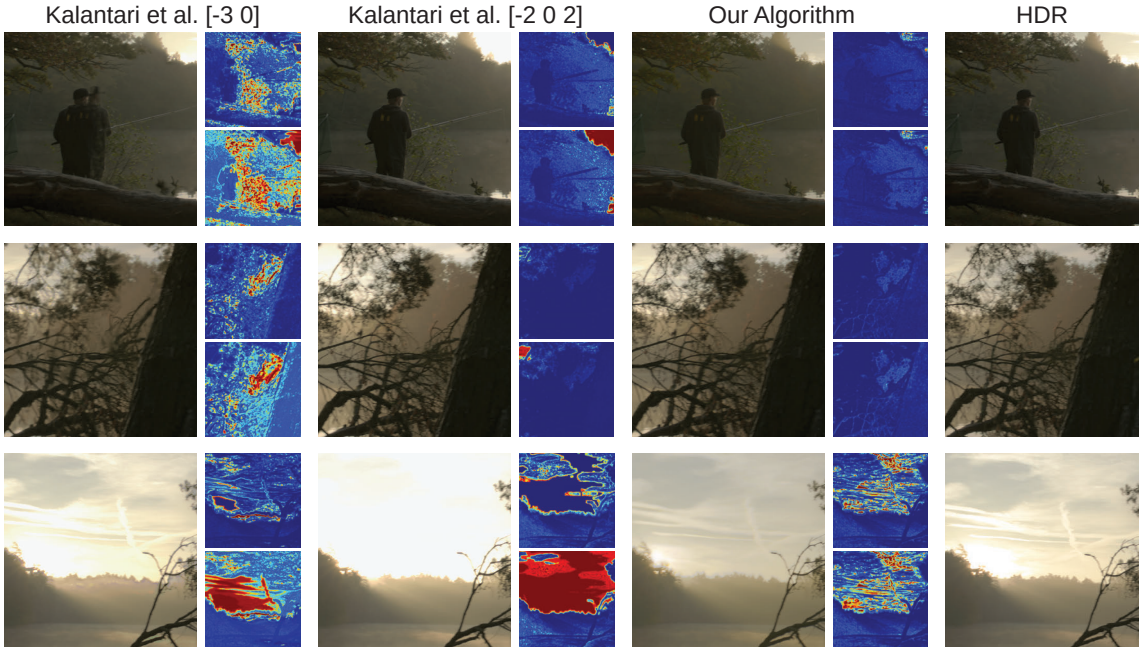


Figure 4.11: Comparisons for three different frames from the Fireplace sequence. See main text for an explanation of the various panels.

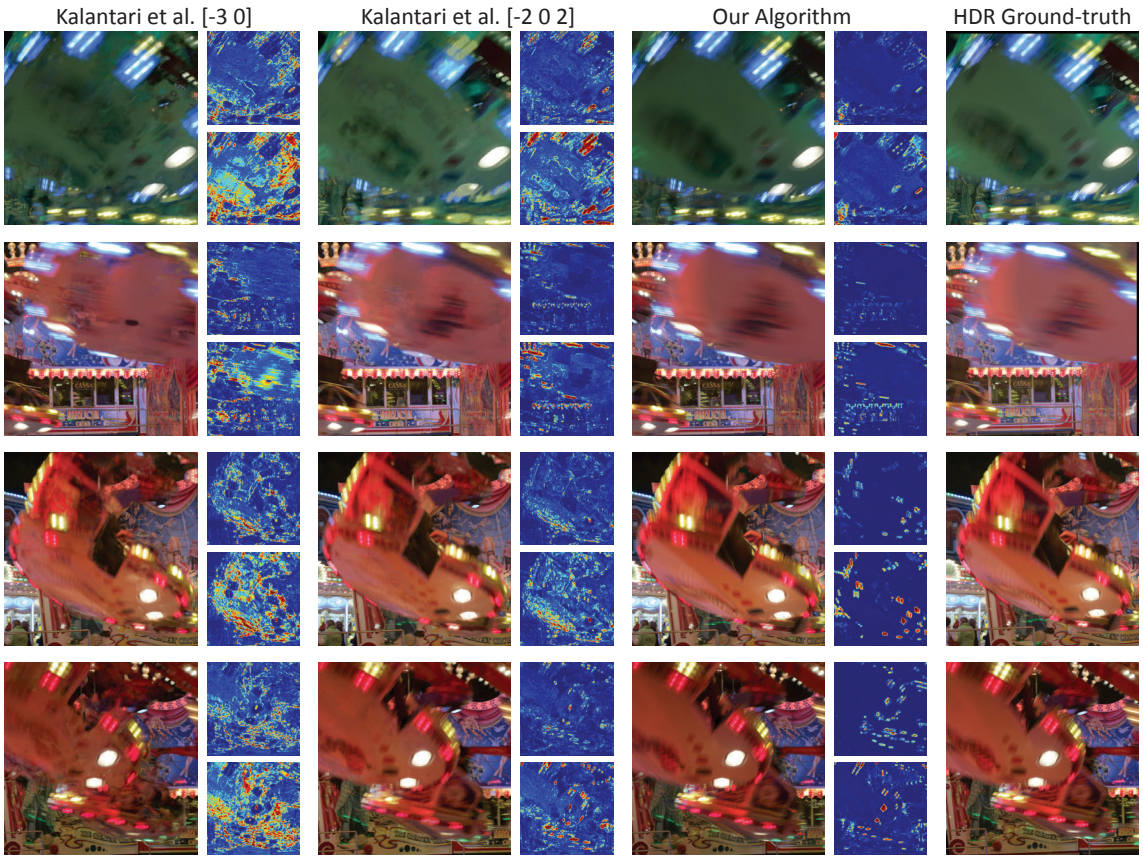


Figure 4.12: Comparisons for four different frames from the Carousel 04 sequence. See main text for an explanation of the various panels.

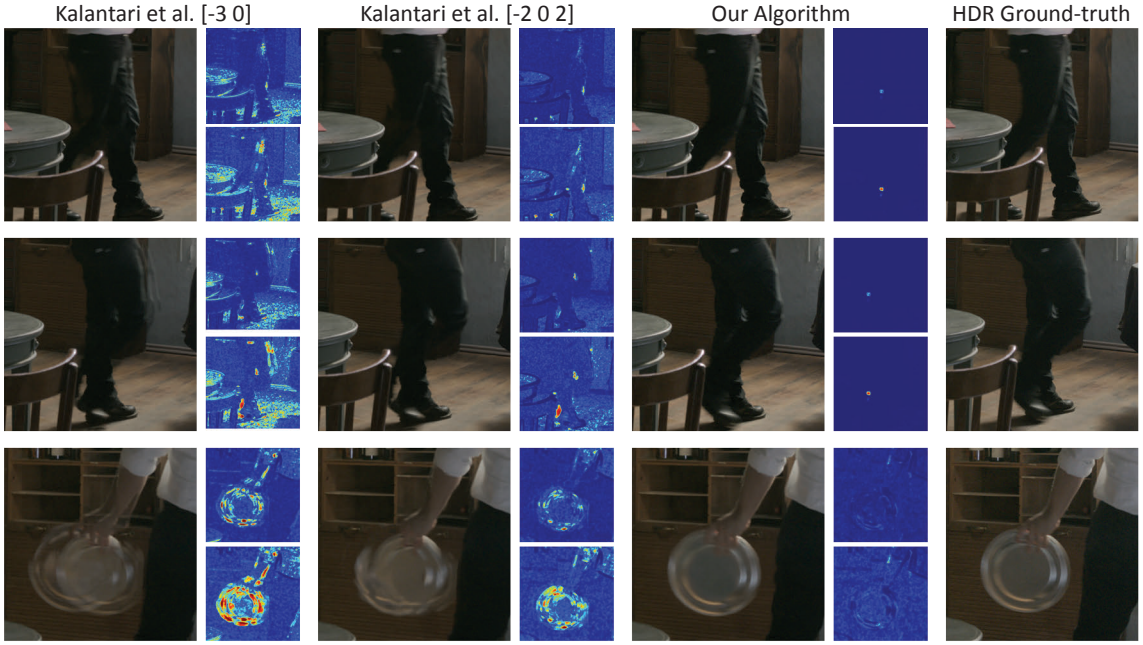


Figure 4.13: Comparisons for three different frames from the Bistro sequence. See main text in this document for an explanation of the various panels.

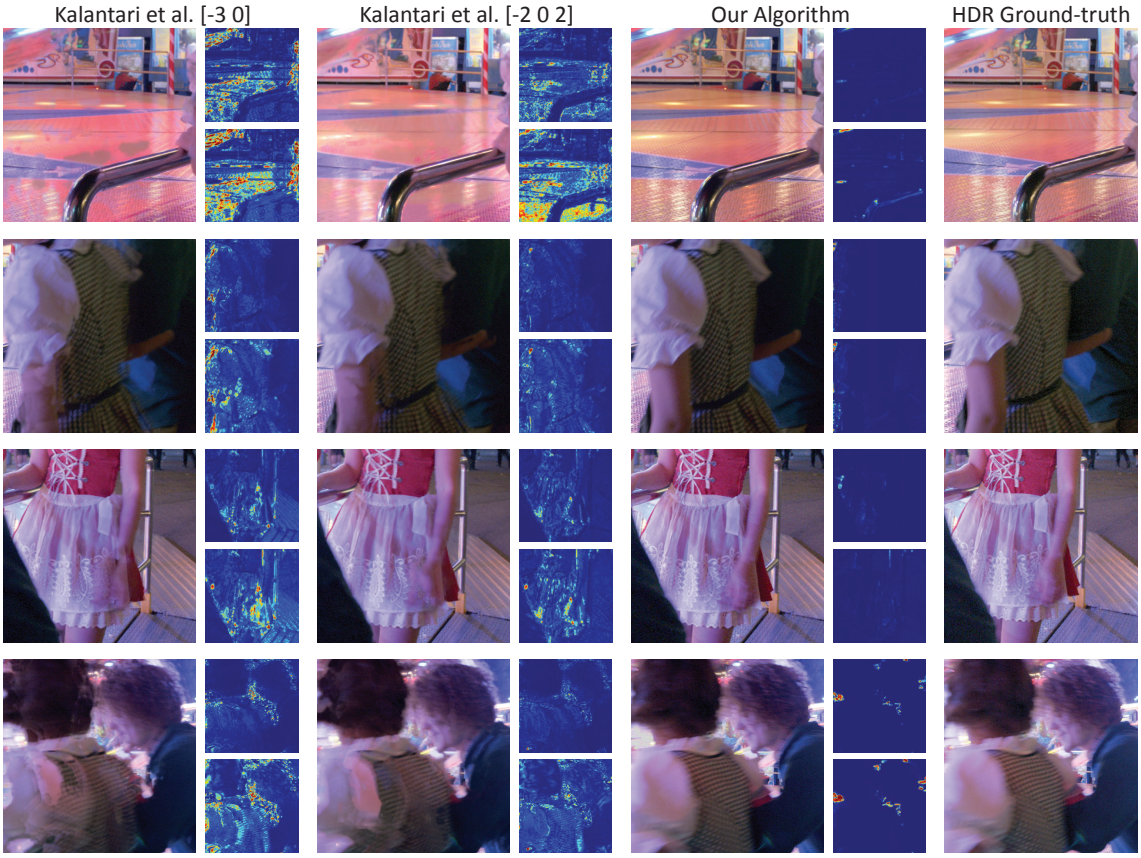


Figure 4.14: Comparisons for four different frames from the Carousel 05 sequence. See main text for an explanation of the various panels.

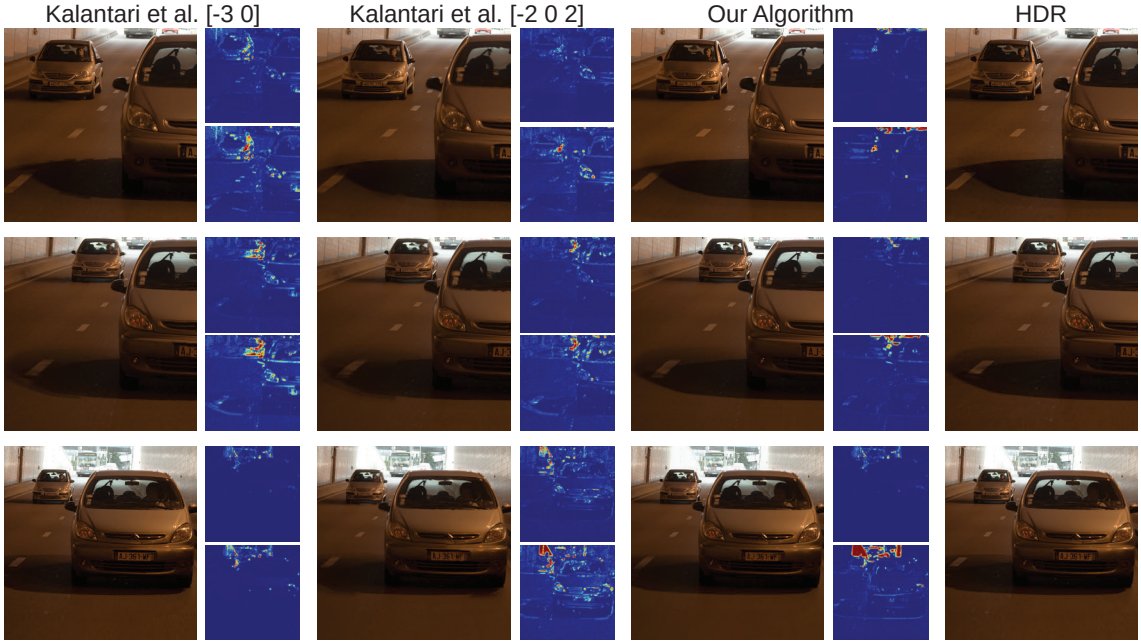


Figure 4.15: Comparisons for four different frames from the Tunnel sequence. See main text for an explanation of the various panels.

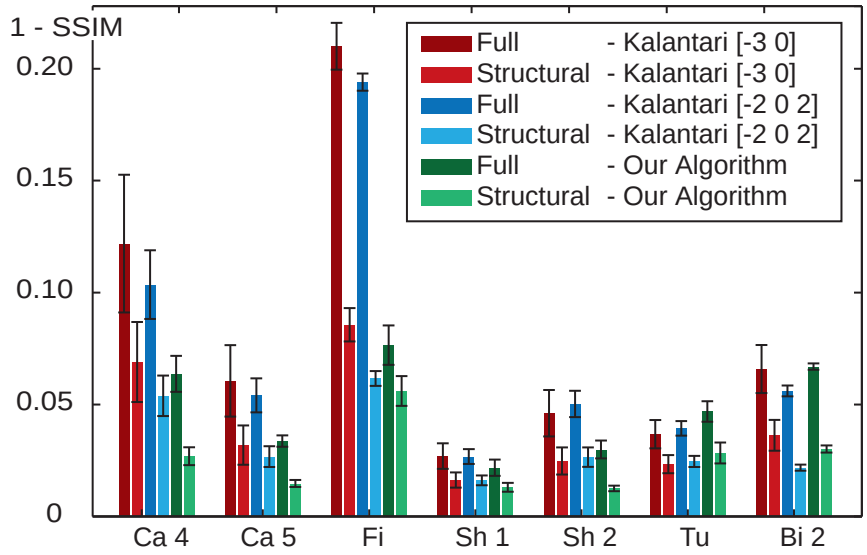


Figure 4.16: Mean SSIM score (full metric and structural component alone) for each of 7 videos, comparing two and three-exposure reconstruction using Kalantari's method with our algorithm. Also the 95% confidence intervals are shown.

indicate that the differences are due to a better performance for our method. Measured per scene, we found our approach to be statistically significantly better for 5 of the 7 scenes, while no significant difference was found for two scenes. To conclude, our metering algorithm adapts well to scenes with varying characteristics, showing a match to the ground-truth data that is consistently no worse than the current state-of-the-art, and in nearly all cases produces visually superior video. This is not to say that our results can never be improved upon. In certain rare cases, we are not able to avoid ghosting completely. For instance, in a high dynamic range scene with significant motion in both the bright and the dark parts, our algorithm will be able to lock onto one of the motion areas, while ignoring the other. This means that one of the two motion areas will be captured well, while the

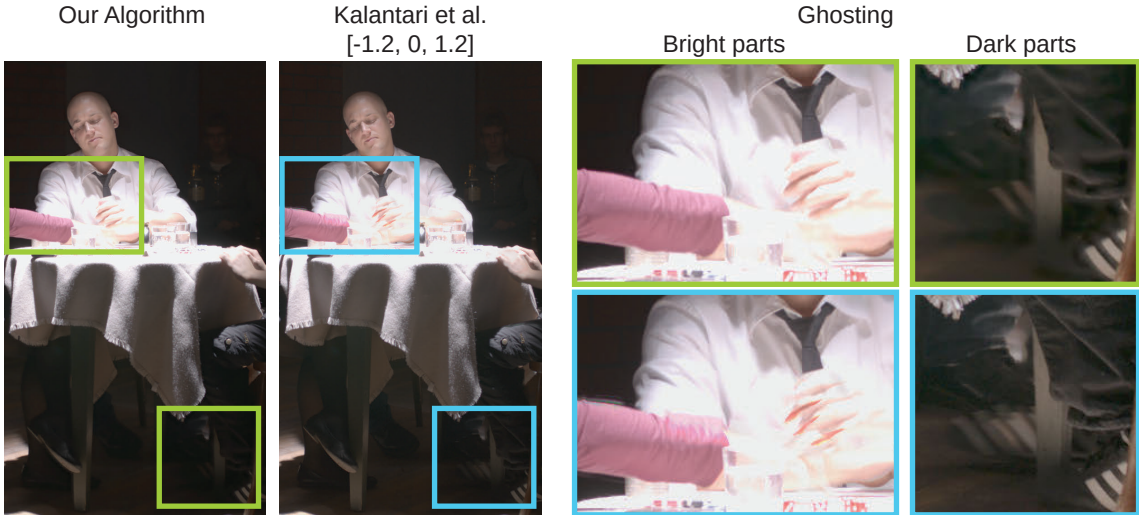


Figure 4.17: Motion in both the bright and the dark parts.

other may exhibit ghosting. We note, however, that competing techniques could produce ghosting in both motion areas (Figure 4.17).

Chapter 5

Sky Based Light Metering for High Dynamic Range Images



a. Sky mask | b. Sky dome (ground-truth / estimated) | c. Tonemapped after scaling with our estimate | d. Tonemapped after scaling with ground-truth | e. CIE ΔE94 between (c) and (d)

Figure 5.1: Our algorithm employs sky models to estimate a scaling that can convert images to absolute luminance values. (a) We detect and segment sky pixels, which are (b) analysed to determine properties of the sky dome and camera parameters (the boxes show the visible area of the sky dome in the image). (c) By scaling the image with our estimation, we obtain absolute radiometric values, allowing images to be processed by appearance models such as Reinhard et al. [2012]. (d) The same processing was applied to the image after scaling it with the ground truth scale factor. (e) The CIE $\Delta E94$ differences between (c) and (d) indicate good correspondence between our estimate and the ground truth.

Image processing and computer vision algorithms often require their input images to have a linear relationship between pixel values and scene luminance. Although knowledge of the camera curve is sufficient for creating such linear content, linearized images can only represent scene luminance up to a constant. Recently, however, several developments have led to an increasing need for image content that is not only linear, but also absolute. This occurs whenever images need to be accurately reproduced on a high dynamic range display, when the image-based recovery of surface reflection parameters is performed or when material characteristics are measured, as well as when interactions between illumination and materials need to be accurately simulated.

Additionally, in image-based lighting [Debevec 2002], light probes captured from real scenes are used to realistically illuminate artificial scenes. Although such light probes are linear, they are not normally specified in absolute units. With absolute light probes, however, image-based lighting becomes a powerful tool for lighting designers or architects. For instance, in the advertisement the rendered cars are often integrated into photographs and videos. The realistic appearance and consistent illumination are achieved by capturing the calibrated HDR environment maps in addition to the target image or video. Further, absolute values are required when one is interested in materials behaviour under different lighting conditions. The use of light probes is beneficial over rendered

skies [Preetham et al. 1999, Hosek and Wilkie 2012] due to variety of sky conditions, which cannot be reproduced by analytic sky-models.

Similarly, absolute scene luminance would be required in image reproduction, and in particular colour appearance modeling. Here, the scene represented by the image would induce a specific state of adaptation in the observer. The display and its environment would induce a different state of adaptation. Colour appearance models account for this mismatch in adaptation [Fairchild 2013] but we can only use such algorithms effectively if we know absolute light levels (Figure 5.2).

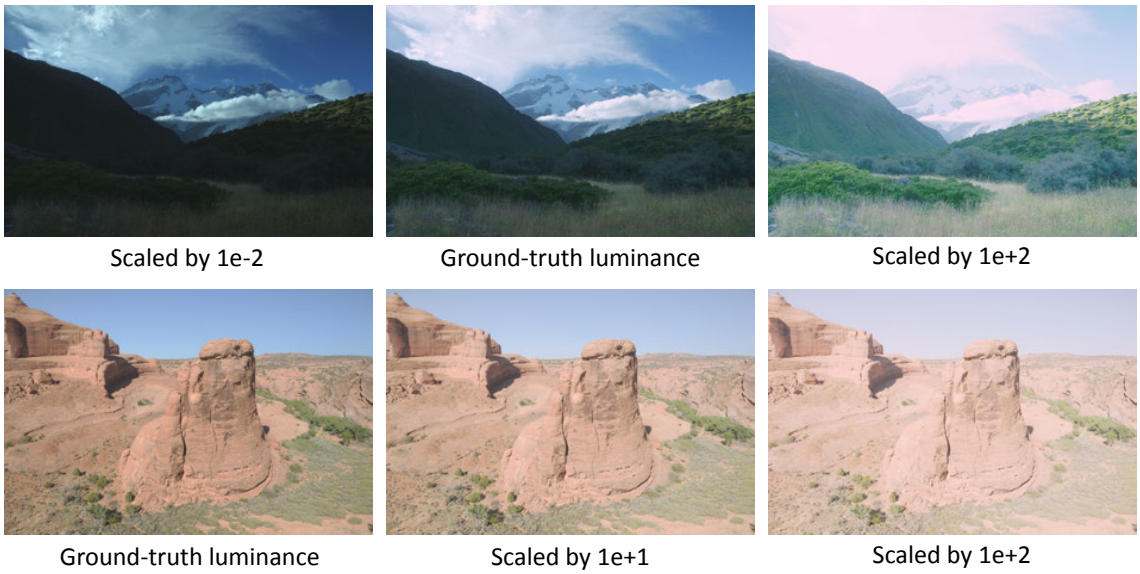


Figure 5.2: An HDR image was scaled to different mean luminances and processed using an appearance model [Reinhard et al. 2012]. The scaled images represent the result depending on the assumed mean luminance of the scene.

A growing number of approaches allows object insertion and relighting from a single image [Karsch et al. 2011, Madsen and Nielsen 2008]. These could benefit from the input to be scaled to absolute values.

While the approach, described in Section 2.4.1, can be used in professional settings when capturing new imagery, the equipment necessary is not universally available and the measurements cannot be taken in retrospect. Consequently, when dealing with existing content (such as light probe galleries used for IBL), only the information within the image and its metadata is available. Moreover, in conditions of fast varying outdoor illumination (clouds, sunset, sunrise, storm etc.) when the photographer is interested in specific lightening phenomena, measurements of the scene are likely to be unreliable or difficult to make.

Without measurements, the absolute values can be obtained under several assumptions about a scene and a camera using EXIF based approaches (Section 2.4.3). In this chapter we discus how the approaches described in Sections 2.4.2 and 2.4.3 can be applied to HDR images.

Further, the possibility of recovering absolute luminance values for outdoor images without the need for professional measurement hardware is explored. The estimation process relies instead on the camera itself as well as a GPS module, which is often included in new camera models.

To validate our algorithm, we require a set of calibrated HDR images with associated EXIF and GPS metadata. To this end, we have acquired a set of ground-truth images, which are publicly

Symbol	Meaning	Symbol	Meaning
θ	Zenith angle	f_c	Focal length of the camera
ϕ	Azimuth angle	ζ	Sun elevation angle
Subscript $_c$	Referring to the camera	t	Sky turbidity
Subscript $_p$	Referring to a sky element in the image	l_z	Relative zenith luminance
Subscript $_s$	Referring to the sun	L_z	Absolute zenith luminance
γ_p	Angle between the sun and a sky element	ξ	Absolute scaling factor
(u_p, v_p)	Coordinates of sky element in image		

Table 5.1: Symbols used in the description of the algorithm.

available [Gryaditskaya et al. 2014]. We show that our algorithm, detailed in Section 5.1, operates with better accuracy than existing approaches (Section 5.4).

5.1 Algorithm

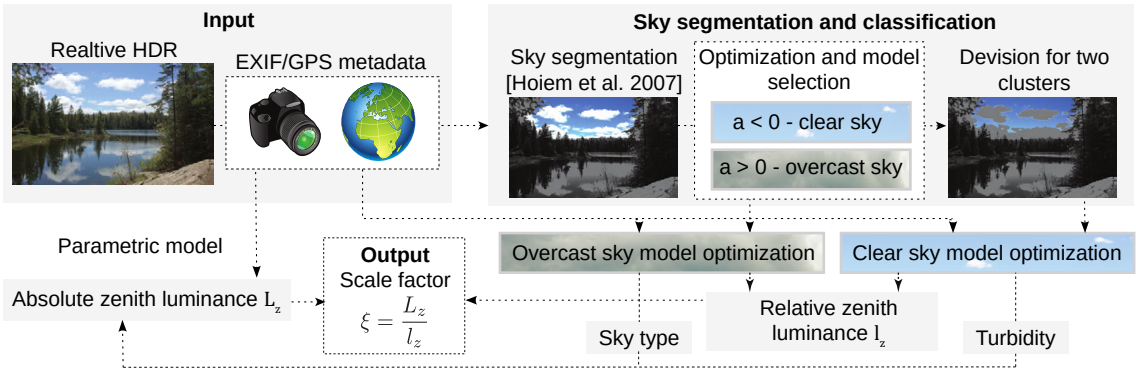


Figure 5.3: The flow of our algorithm. Note that when the input is an environment map the GPS/EXIF data are not required.

The goal of our algorithm is to convert a linear HDR image of an outdoor scene into absolute radiometric measurements. To achieve that we compute a scale factor ξ, which converts relative luminance values to absolute luminance. The input to our algorithm is a single linear HDR image, GPS data and a small set of metadata, which can be extracted from the EXIF data of the image itself (date and time of capture, sensor size, focal length). If an environment map is used no GPS or EXIF information is required. Figure 5.3 illustrates the flow of the algorithm.

To compute ξ, we rely on sky regions in the image, found using a semantic image segmentation algorithm [Hoiem et al. 2007]. Unlike most parts of an outdoor scene, the appearance of the sky is predictable and can be modelled with reasonable accuracy given a small set of parameters. This

observation forms the main motivation for our framework; since pixels within an HDR image are linearly related, it is sufficient to estimate a scaling factor for sky pixels in the image.

Our algorithm first estimates the zenith luminance l_z of the sky dome relative to the sky pixels visible in the image. This is obtained by fitting the intensity values of the observed pixels to a parametric sky model [Perez et al. 1993] and using the zenith value as a normalization factor to the range of the values in the image. During the fitting procedure the unknown parameters for the sky model are estimated as well as l_z. Using information obtained during this step, as well as some of the information within the metadata, we can also obtain an absolute estimate of the zenith luminance L_z. Since the original zenith estimate is given relative to the values of sky pixels in the image, this second absolute estimate allows us to compute a scale factor, that can finally be applied to all pixels:

$$\xi \triangleq L/l = L_z/l_z, \tag{5.1}$$

where L denotes absolute and l relative luminance values.

5.1.1 Sky Appearance

When sunlight enters the atmosphere, it interacts with particles and molecules before reaching ground level. Small particles cause Rayleigh scattering (which leads to blue skies and a yellow sun), whereas larger molecules create Mie scattering, which is not strongly wavelength dependent but leads to haze. The latter is often described with a single parameter, called turbidity [McCartney 1976]. The luminance distribution of the sky dome thus depends on atmospheric conditions as well as the position of the sun. Various models exist that predict this distribution. In our work, we use the Perez sky model as a prior, which was shown to out-perform competing algorithms [Ineichen and Molineaux 1994]. It allows us to infer the luminance arriving from the zenith, based on the sky region visible in the input HDR image.

The Relative All-Weather Perez Sky Model

The model introduced by Perez et al. [1993] consists of two factors. The first factor describes vertical gradations of luminance values and is therefore called the gradation function:

$$f_{\text{grad}}(\theta_p, a, b) = 1 + a\exp(b/\cos\theta_p), \tag{5.2}$$

where θ_p is the zenith angle of the considered sky element (see Figure 5.4), a and b are two parameters which describe sky properties. The first parameter a corresponds to darkened ($a > 0$) or lightened horizon regions ($a < 0$), which is indicative of whether the sky is overcast or not. The second parameter $b \in [-1, 0]$ refers to changes of the luminance gradient near horizon. This factor describes sky luminances away from the sun.

The second factor describes the influence of the sun on the sky and is given by the indicatrix function:

$$f_{\text{ind}}(\gamma_p, c, d, e) = 1 + c\exp(d\gamma_p) + e\cos^2\gamma_p, \tag{5.3}$$

where γ_p is the angle between a sky element and the position of the sun (Figure 5.4) and c, e, d are parameters describing the sky appearance near the sun. Parameter $c \geq 0$ is proportional to the luminance of the circumsolar region, while d is its width. Finally, e accounts for the relative

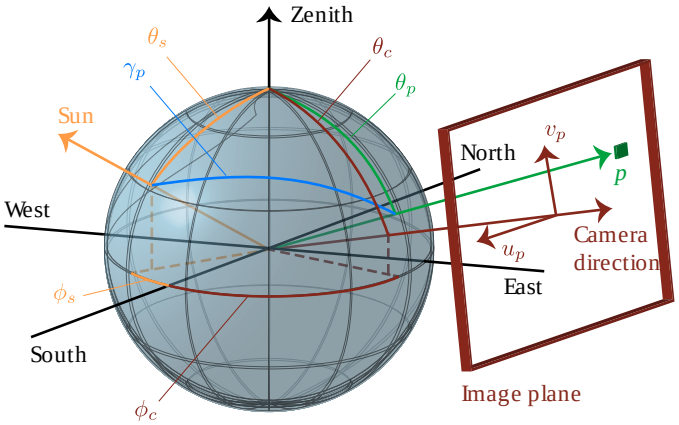


Figure 5.4: Angles defining the position of the sun, a camera and a sky pixel p in the image.

intensity of backscattered light. Combining these parameters into a vector $\vec{\rho} = (a,b,c,d,e)$, Perez's full model is then the product of these two factors:

$$f(\theta_p,\gamma_p,\vec{\rho}) = f_{\text{grad}}(\theta_p,a,b) f_{\text{ind}}(\gamma_p,c,d,e) \tag{5.4}$$

The values returned by the Perez sky model are not absolute and can be normalized to any desirable range of values, using a known luminance value of any sky element. Typically, the zenith luminance l_z is used:

$$l_p = l_z \frac{f(\theta_p,\gamma_p,\vec{\rho})}{f(0,\theta_s,\vec{\rho})}, \tag{5.5}$$

where $f(0,\theta_s,\vec{\rho})$ is the relative luminance at the zenith, predicted by (5.4), as in this case the zenith angle θ_p of the sky element is equal to zero, and the angle of the sky element with the sun γ_p is equal to the sun zenith angle θ_s (see Figure 5.4).

The five parameters in Perez's model directly relate to turbidity t, especially for relatively low turbidities ($t \in [2,6]$) that describe atmospheres that are between clear and thin fog [Preetham et al. 1999];

$$\vec{\rho}^T = \begin{bmatrix} t \\ 1 \end{bmatrix}^T \begin{bmatrix} 0.179 & -0.355 & -0.023 & 0.121 & -0.067 \\ -1.463 & 0.428 & 5.325 & -2.577 & 0.370 \end{bmatrix} \tag{5.6}$$

Given this relation, we will use $f(\theta_p,\gamma_p,\rho)$ and $f(\theta_p,\gamma_p,t)$ interchangeably, but note that turbidity only describes clear to hazy skies; it is not appropriate for overcast skies.

Overcast skies can be classified into one of five general types [Kittler et al. 1998]. These types are described in terms of the five sky parameters a,b,c,d,e from Perez's sky model [Perez et al. 1993]. The parameters a and b define the representation of gradation function. Here, the types of gradation function are denoted by Roman numerals. The parameters c, d and e in turn define the representation of the indikatrix function, whose types are denoted by Arabic numbers. The values of the five parameters describing each of five sky types can be found in Table 5.2 and the relative luminance distribution for each sky type is shown in Figure 5.5. The description of these five sky types is as follows:

I.1 Sky with steep luminance gradation towards zenith and azimuthal uniformity

Type	Gradation	Indikatrix	a	b	c	d	e	B	C	D	E	D_v/E_v
1	II	2	1.1	-0.8	2	-1.5	0.15	12.23	3.57	0.57	44.27	0.22
2	I	2	4.0	-0.7	2	-1.5	0.15	12.35	3.68	0.59	50.47	0.18
3	III	1	0.0	-1.0	0	-1.0	0.00	42.59	1.00	0.00	0.00	0.20
4	II	1	1.1	-0.8	0	-1.0	0.00	48.30	1.00	0.00	0.00	0.10
5	I	1	4.0	-0.7	0	-1.0	0.00	54.63	1.00	0.00	0.00	0.10

Table 5.2: Typical parameters for five standard overcast sky types.

II.1 Sky moderately graded with azimuthal uniformity
III.1 Sky of uniform luminance
I.2 Sky with steep luminance gradation towards zenith and slight brightening towards the sun
II.2 Sky moderately graded with slight brightening towards the sun

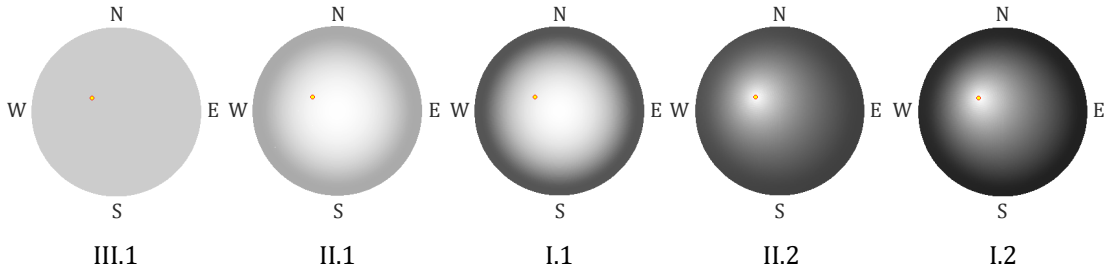


Figure 5.5: Relative distribution of luminance over the sky dome for each of the five overcast sky types. The yellow circle denotes sun position.

5.1.2 The Fitting to Parametric Sky Model

The key idea of our approach is to extract information from the image by fitting a parametric sky model to the intensity values of the observed sky region. In previous works, sky regions were used to estimate camera spectral sensitivity and white balancing [Kawakami et al. 2013]. Using time-lapse sequences, skies were also used as geometric calibration targets to reverse-engineer intrinsic camera and lens parameters, including the focal length of the lens, the orientation of the camera as well as its geolocation [Lalonde et al. 2010]. In a sense, our work is complementary to this analysis.

To compare image intensities with the sky luminance distribution two operations should be performed. First, the relation of pixel in position (u_p, v_p) in the image and the corresponding sky element p in the sky dome should be set (see Figure 5.4). This is obtained by expressing the sky element zenith and azimuth angles in terms of the camera parameters and pixel coordinates. The

zenith angle θ_p of the sky element and its angle with the sun γ_p are [Lalonde et al. 2010]:

$$\theta_p = \cos^{-1}\left(\frac{v_p \sin\theta_c + f_c \cos\theta_c}{\sqrt{f_c^2 + u_p^2 + v_p^2}}\right) \tag{5.7}$$

$$\gamma_p = \cos^{-1}\left(\cos\theta_s \cos\theta_p + \sin\theta_s \sin\theta_p \cos(\phi_p - \phi_s)\right), \tag{5.8}$$

where ϕ_s is the sun azimuth angle and the sky element azimuth angle ϕ_p is found from the equation:

$$\tan(\phi_p) = \frac{f_c \sin\phi_c \sin\theta_c - u_p \cos\phi_c - v_p \sin\theta_c \cos\phi_c}{f_c \cos\phi_c \sin\theta_c + u_p \sin\phi_c - v_p \cos\theta_c \cos\phi_c}.$$

Now, Equation (5.4) can be expressed in terms of camera parameters and pixel coordinates:

$$f(\theta_p, \gamma_p, \rho) = g(u_p, v_p, \theta_c, \phi_c, f_c, \theta_s, \phi_s, \rho), \tag{5.9}$$

Second, the luminance values produced by g should be normalized to the range of the intensities in the image. For this, we use Equation (5.5).

The fitting problem can then be formulated as an optimization (see Table 6.1):

$$\min_{\theta_c, \phi_c, \vec{\rho}, l_z} \sum_{p \in P} \left(l_p - l_z \frac{g(u_p, v_p, \theta_c, \phi_c, \vec{\rho})}{f(0, \theta_s, \vec{\rho})}\right)^2, \tag{5.10}$$

where P is the set of sky pixels and l_p is the observed relative intensity of pixel p in the image. Here, the nominator produces the zenith luminance value predicted by the sky model, and serves as a normalization term.

The atmospheric conditions are specified in the Perez sky model using five parameters $\vec{\rho}$. These parameters determine the appearance of the sky and can be related to turbidity t, which is a single parameter, commonly used to describe how clear or foggy the sky is. For low turbidities ($t \in [2,6]$), i.e. between clear sky and thin fog [Preetham et al. 1999], the parameters $\vec{\rho}$ can be mapped to turbidity t directly, allowing us to optimize over one parameter t, rather than five parameters $\vec{\rho}$.

To further reduce the number of unknowns, we take advantage of the increased availability of GPS systems and the EXIF data that is typically attached to the image. Using GPS data, time information from EXIF and a known solar model [Reda and Andreas 2004] we evaluate the sun zenith θ_s and azimuth ϕ_s angles. The image's EXIF data is used to compute the focal length in pixels, which is obtained from the focal length (usually given in *mm*) and the size of the camera sensor. Having the focal length allows us to make the optimization quite robust and makes the algorithm sufficiently stable to rely on the information obtained from a single image in comparison with images sequence [Lalonde et al. 2010].

To minimize the risk of convergence to a local minimum we impose constraints on the sky appearance parameters $\vec{\rho}$ or t, which is possible due to their limited range of values [Perez et al. 1993, Preetham et al. 1999]. Following Lalonde et al. [2010] we initialize the camera zenith angle θ_c from the assumption that the lowest sky pixel corresponds to the horizon. We also initialize the camera azimuth angle ϕ_c with a set of values in the range $[0, 2\pi]$ and choose the solution that gives the smallest residual. Note that especially when the camera is facing away from the sun, the accuracy of this solution is non-crucial (see Figure 5.1b).

Being a non-linear least-squares minimization problem with inequality constraints, (5.10) requires an iterative approach for its solution. We use the implementation of the trust-region-reflective algorithm from the MATLAB Optimization Toolbox as it fully satisfies our requirements.

5.1.3 Sky Segmentation and Classification

The first step of our algorithm is to extract portions of the image that depict sky. We segment the image using the segmentation method of Felzenszwalb and Huttenlocher [2004] and use the surface layout recovery method of Hoiem et al. [2007] to separate sky regions from the remainder of the image. Alternatively, the method of Tao et al. [2009] can be used to segment the sky region.

We wish to determine if this sky segment represents a clear or overcast sky. The segment's pixel data is conditioned by excluding 20% of the lightest pixels, as regions near the sun tend to be unreliable due to exposure problems. We note that one of parameters $a \in \vec{\rho}$ of the Perez sky model is indicative of whether sky pixels near the horizon are lighter than sky pixels away from the horizon ($a < 0$, indicating clear sky), or vice-versa ($a > 0$, indicating overcast sky). Therefore solving (5.10) for this parameter allows us to automatically determine whether the sky is clear or overcast and choose the optimal algorithm.

The necessity of having two algorithms is explained by the differences in absolute zenith luminance models, which we use for the scale factor estimation (5.1). For our datasets, we find that this approach correctly selects the type of sky for 95% of our clear images as well as for more than 80% of our overcast images. In the case of clear sky images, the segmentation can be further fine-tuned using k-means clustering, whereby the cluster closer to blue is selected.

5.1.4 Relative Zenith Luminance

We are interested in recovering the zenith luminance l_z relative to the luminance values in the image. One can think about the value of l_z as the value that would be recorded if the image could be extended so that the zenith was visible.

Images with clear sky regions. The relative zenith luminance l_z for the case of clear skies is found as a solution of (5.10), where instead of optimization by $\vec{\rho}$ the optimization is performed by turbidity t.

Images with overcast skies. Since the appearance of overcast skies is significantly different from clear skies, a different approach is required. The relative zenith luminance l_z estimation for overcast images is performed in two steps:

I. First, we solve a continuous optimization problem (5.10) to obtain an initial estimate of the model parameters.

We consider several alternative approaches for the second step:

II.a We estimate the type of overcast sky i_{opt} in the image following Kittler's classification for five types of overcast skies [Kittler et al. 1998] (Figure 5.5). The sky type is required for absolute zenith luminance estimation. Let $\vec{\rho}_i$ be the set of sky parameters of one of the five overcast sky types, $i \in [1,5]$. Then, the sky type i_{opt} can be taken as the one with the lowest l^2-norm of the differences between its vector of sky parameters $\vec{\rho}_{i_{\text{opt}}}$ and the estimated optimal vector $\vec{\rho}_{\text{est}}$ for a given image. Afterwards, the problem (5.10) can be solved again to find a better estimate of the relative zenith luminance l_z, with $\vec{\rho}$ set to the values of sky parameters $\vec{\rho}_{i_{\text{opt}}}$ of the found sky type.

II.b Instead of comparing the optimal vector ρ_{est} of sky parameters with a vector of parameters describing certain type of a sky, we can take the camera zenith $\theta_{c_{\text{est}}}$ and azimuth $\phi_{c_{\text{est}}}$ angles as exact. Then, for each sky type i the problem formulated on the first step of the algorithm can be solved with θ_c, ϕ_c and ρ set to $\theta_{c_{\text{est}}}, \phi_{c_{\text{est}}}$ and ρ_i. The sky type i_{opt} can then be determined as one

Model	Description	k_0	$10^2\,k_1$	$10^3\,k_2$	$10^4\,k_3$	$10^6\,k_4$	$10^9\,k_5$
Soler13_1	Zenith luminance 15–min means	-0.68260	9.922	-5.16	1.44300	1.8918	9.3644
Soler13_2	Zenith luminance at every 5° interval of ζ	-0.60647	8.594	-4.44	1.29610	1.7942	9.4016
$\text{Soler}_{t<3}$	Turbitidy $t < 3$	-0.61538	8.580	-4.48	1.26190	1.6824	8.54789
$\text{Soler}_{3<t<5}$	Turbidity $3 < t < 5$	-0.60879	9.129	-4.52	1.22350	1.5829	7.84640
$\text{Soler}_{t>5}$	Turbidity $t > 5$	-0.58487	0.023	-3.88	0.99654	1.2851	6.55570

Table 5.3: Coefficients for 5th degree polynomial zenith luminance models from Soler and Gopinathan [2000].

for which the value of the objective function of the problem from the first step in the solution is minimal.

II.c In some cases, specifically when the camera is facing away from the sun, only the zenith angle of the camera θ_c plays an important role in the calculations while errors in the azimuth angle ϕ_c do not affect the accuracy of the result. To cover this scenario, we also consider the case when $\theta_{c_{\text{est}}}$ is assumed to be an exact value of the zenith angle. In this case, we fit the value of the camera azimuth angle ϕ_c in the process of solving the indicated problem for each sky type i.

5.1.5 Absolute Zenith Luminance

To scale the image to absolute radiometric quantities, we also need an estimate of absolute luminance. There exist several models which correlate the luminance arriving from the zenith to the solar elevation as well as turbidity for cloudless skies [Karayel et al. 1984, Preetham et al. 1999, Soler and Gopinathan 2000] as well as for overcast skies [Soler et al. 2001]. The time of year and the luminous clearness index of the sky is also indicative of zenith luminance [Soler and Gopinathan 2001].

For clear skies, we selected zenith luminance models from Karayel et al. [1984] and Soler and Gopinathan [2000], as these were found to perform well in a comparative study [Zotti et al. 2007]. Zenith luminance L_z^{Karayel} can be computed from turbidity t and sun elevation angle ζ [Karayel et al. 1984]:

$$L_z^{\text{Karayel}} = (1.376t - 1.81)\tan(\zeta) + 0.38, \tag{5.11}$$

where t is turbidity and ζ is the sun elevation angle. Soler and Gopinathan [2000] propose to fit fifth order polynomials to measured data as function of sun elevation angle ζ:

$$L_z^{\text{Soler}} = \exp\left(\sum_{i=0}^{5} k_i \zeta^i\right) \tag{5.12}$$

where k_i are coefficients. They derived a set of five models for various different conditions. The parameters for each of these models are given in Table 5.3.

We found that while both options give reasonable results, their accuracy depends on the sun elevation angle. To improve the accuracy over either model, we propose a weighted combination of

L_z^{Karayel} and L_z^{Soler}, interpolating according to sun elevation angle $\zeta \in [0,75]$ (degrees):

$$L_z = \cos^2(3\zeta) L_z^{\text{Karayel}} + \sin^2(3\zeta) L_z^{\text{Soler}}. \tag{5.13}$$

Based on the above interpolation, Soler and Copinathan models [2000] are prioritized for values between 15 and 45 degrees, whereas model of Karayel et al. receives a higher weight for all other angles. Note that the interpolation weights always sum to 1. We validate this choice in Section 5.4.

For overcast skies, models of zenith luminance can be derived for different sky types ranging from completely overcast to near overcast [Kittler et al. 1998, Soler et al. 2001]. In particular, Kittler et al. [1998] distinguish five different types of overcast skies. Zenith luminance of a given sky type can be found using Table 5.2 [CIE 2003]:

$$L_z^{\text{overcast}} = \frac{D_v}{E_v}\left(\frac{B(\sin(\zeta))^C}{(\cos(\zeta))^C} + E\sin(\zeta)\right), \tag{5.14}$$

where D_v is diffuse sky illuminance, E_v is extraterrestrial horizontal illuminance, ζ, as previously, is the sun elevation angle, and B, C and E are parameters determined by the given sky type.

5.2 Scale Factor Estimation for Angular Maps

In the introduction of this chapter we discuss how environment maps depicting the sky are used to achieve realistic daylighting. In this section, we describe how our algorithm can be used to scale images of this form to absolute values.

The sky maps depict the whole sky hemisphere. Thus, the zenith sky element is visible in these images, potentially making the problem simpler. However, even if the sky has clear regions the sky zenith element could still be occluded by clouds. For overcast skies the visible sky element could not sufficiently represent the whole sky appearance. So, the optimization by the relative zenith luminance l_z could still be beneficial.

To perform the optimization we first need to express zenith angle θ and azimuth angle ϕ as functions of pixel coordinates. Here we give expressions for angular and latitude-longitude map formats. The conversions between different environment map formats can be found in [Reinhard et al. 2010].

Beginning with an angular map with pixel coordinates (u,v) normalised to the $[-1,1]$ range, we derive polar coordinates $(\alpha,\beta) = (\pi\sqrt{u^2+v^2}, \tan^{-1}(v/u))$. This allows us to express zenith and azimuth angles for each pixel as follows:

$$\theta_p = \cos^{-1}(\sin(\alpha)\sin(\beta)) \tag{5.15}$$

$$\phi_p = \tan^{-1}(\sin(\alpha)\cos(\beta)/\cos(\alpha)). \tag{5.16}$$

For latitude-longitude maps the zenith and azimuth angles can be evaluated directly from normalized pixel coordinates with $(\theta_p,\phi_p) = (\pi v/2, \pi u)$.

Further, we need to find the sun position. If the sun is not occluded its brightness typically exceeds the brightness of sky and clouds by a factor of fifty thousand. Correct maps should preserve this ratio. The technique of capturing bright sources is described in detailed by Stumpfel et al. [2004].

This ratio allows us to estimate the sun position as the point with the maximum intensity value in the image. Even if the sun is occluded by clouds this assumption will either hold or the environment will correspond to an overcast sky type where the sky appearance doesn't depend on the sun position.

Then, sun zenith and sun azimuth angles could be found by substituting sun coordinates in the light probe into the above equations.

Estimating the sun position from the image itself allows us to lift the requirement of having GPS coordinates and time of image capture. Additionally, expressing the zenith and azimuth angles as functions of pixel coordinates only allows us to lift the requirement of EXIF data as well. Moreover, we do not require the camera direction. The visibility of the whole sky hemisphere allows us to obtain the angle γ_p between sun direction and sky element p's direction from an image without knowledge of their absolute positions in the world. This enables the elimination of the camera zenith angle from the optimization. Thus, relative zenith luminance l_z can be estimated as the solution to the following problem:

$$\min_{t,l_z} \sum_{p \in P} \left(l_p - \frac{l_z}{f(0,\theta_s)} f(\theta_p, \gamma_p, t) \right)^2 . \tag{5.17}$$

The absolute zenith luminance and the final scale factor can be found as described before.

5.3 Ground Truth Dataset

To validate our algorithm, we require a calibrated set of HDR images that are annotated with both EXIF and GPS data. To our knowledge, only the HDR Survey [Fairchild 2007] provides all this information. As this set contains a limited number of images with sky regions, we have created an additional calibrated HDR image dataset, annotated with EXIF as well as GPS location and orientation information. All images in this dataset contain a significant portion of clear or overcast sky.

Our HDR dataset was photographed with a tripod-mounted Nikon D2h digital SLR camera, which acquires up to 9 images each spaced 1 EV apart, using autobracketing. The white point was set to a fixed 6700K (nearest to D65 that this camera supports) and the exposures were saved in the sRGB colour space. Images were assembled into linear HDRs using Greg Ward's Photosphere application, which we have also used to derive the response curve of the Nikon D2h. We placed both an 18% grey card and a GretagMacbeth colour checker in each scene for calibration. To scale the HDR images to absolute values, measurements of Yxy components of the grey card and several patches from the ColorChecker were taken with a Minolta CS100 colour and luminance meter. We also transformed the images to the D65 white point, as per the sRGB standard. Using this set-up we collected 25 calibrated HDR images containing clear skies with varying turbidity and 9 images with overcast skies.

5.4 Algorithm Validation

In the following, we compare variants of our model for clear and overcast skies on a selection of images from the HDR Photographic Survey that contain sufficiently large areas of skies as well as on the images from our dataset. We use relative error in our comparisons: $\delta \triangleq |\xi - \xi_{\text{meas}}|/\xi_{\text{meas}}$, where ξ_{meas} is the scale factor taken from the HDR Photographic Survey or measured as described in Section 5.3.

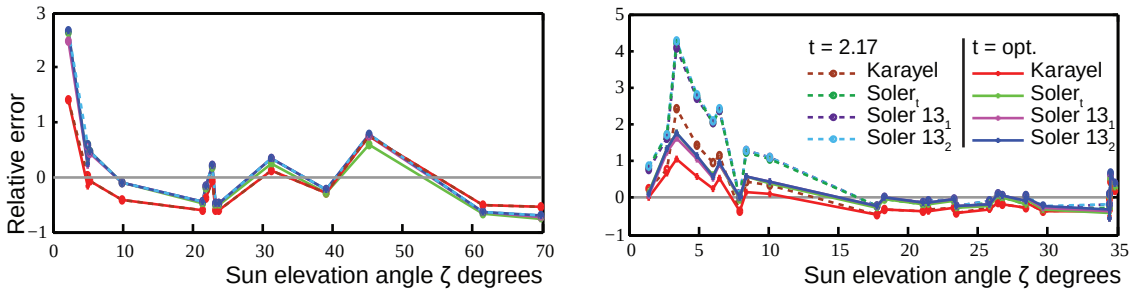
Zenith	$\mu(\lvert\delta\rvert)$			
Lum.	$t = 2.17$	t optimized		
Model	all ζ	all ζ	$\zeta < 15$	$\zeta > 15$
Karayel	0.459	0.469	0.515	0.450
Soler_t	0.586	0.563	0.863	0.442
Soler13_1	0.577	0.554	0.815	0.449
Soler13_2	0.795	0.572	0.875	0.450

Table 5.4: Mean μ of relative errors magnitudes $|\delta|$ for images from the HDR Photographic Survey.

5.4.1 Clear-Sky Images

The scale factor that allows the creation of absolute pixel values, is estimated using the various absolute zenith models (Section 5.1.5) to calculate L_z. The results for different types of Soler and Gopinathan [2000] zenith luminance models are shown in Figure 5.6, and aggregated in Tables 5.4 and 5.5. We further find that including turbidity t in the optimization provides a significant advantage over fixing turbidity to the constant value of $t = 2.17$, which is the CIE-defined turbidity for a clear sky [CIE 2003]. It could be seen that our optimization algorithm is sufficiently stable to include turbidity in the calculations.

The mean and standard deviation of errors obtained using our weighted combination of models proposed in [Karayel et al. 1984] and [Soler and Gopinathan 2000] are listed in Table 5.6. As all three Soler models produce comparably good results, we chose the model $Soler_t$ because of its lower standard deviation.



a. HDR Survey b. Our images

Figure 5.6: Relative errors for images with clear sky regions from the HDR Survey [Fairchild 2007] (left) and from our own dataset (right).

We note that for sun elevation angles below 15 degrees Karayel's model performs best, whereas Soler's models perform better for sun elevations between 15 and 50 degrees. Our weighted combination of both models (Eq. (5.13)) accounts for this and produces errors smaller than can be achieved with either model. Using this weighted combination we obtain a mean error of 0.440 $\pm$ 0.344 (s.d.) for the HDR photographic survey [Fairchild 2007] and 0.309 $\pm$ 0.251 for our own dataset.

Zenith	$\mu(\lvert\delta\rvert)$			
Lum.	$t = 2.17$	Opt. t		
Model	all ζ	all ζ	$\zeta < 15$	$\zeta > 15$
Karayel	0.503	0.354	0.416	0.319
Soler$_t$	0.759	0.416	0.769	0.218
Soler13$_1$	0.745	0.396	0.724	0.211
Soler13$_2$	0.773	0.413	0.768	0.212

Table 5.5: Mean values μ of relative error magnitudes $|\delta|$ in our clear-sky image collection.

Algorithm	HDR Survey		Our data	
	$\mu(\lvert\delta\rvert)$	$\sigma(\lvert\delta\rvert)$	$\mu(\lvert\delta\rvert)$	$\sigma(\lvert\delta\rvert)$
Soler$_t$	0.440	0.334	0.309	0.251
Soler13$_1$	0.445	0.345	0.305	0.255
Soler13$_2$	0.445	0.346	0.307	0.256

Table 5.6: Mean μ and standard deviation σ of relative error magnitudes $|\delta|$ for images with clear sky regions, when L_z is computed with one of set of parameters listed in Table 5.3.

Note that though the measurements for indoor scenes can be taken with high precision, it is not an easy task for outdoor images. Thus, the measured values for some images in the HDR photographic survey could give scale factors different from each other by 40%. Additionally, for a sufficient number of images from Fairchild's dataset the scale factor is estimated from measurements of a scene itself, which gives additional imprecision in ground-truth scale factor estimation. This can explain the higher mean relative error for images from the HDR photographic survey.

To demonstrate the performance of our approach we also compare our algorithm with techniques that solely rely on EXIF data. Such techniques compute the mean luminance of a scene as follows:

$$L = K\frac{N^2}{St}, \tag{5.18}$$

where K is a constant, N is the f-number, S is sensitivity and t is the exposure time. In the general case these techniques rely on several assumptions. The first is that the mean luminance in the area measured by the in-camera light meter corresponds to the luminance of an 18% card if this card would be inserted into the scene and measured. The second pertains to the used definition of sensitivity. The ISO standard 12232 [ISO 2006] includes several definitions of sensitivity: saturation-based ISO sensitivity and standard output sensitivity (see supp. materials). As discussed in [Ackermann and Goesele 2013] using the different definitions of sensitivity will give significantly different estimations for the same pixel. In their example they get two estimations, different by a factor of 1.43.



Figure 5.7: Two exposure sequences for HDR images of the same scene. The first scene is exposed for the background and the second one for the person standing behind. The exposure compensation value of the middle image in both sequences is zero.

Image	Mean luminance		
	EXIF	Our algorithm	Measured
background	12.5e+03	2.41e+03	2.95e+03
foreground	1.0e+03	2.59e+03	

Table 5.7: Mean luminance values for the images obtained from HDR sequences in the Figure 5.7, estimated using the EXIF-based approach (5.18) or by scaling mean luminance of an HDR image by a scale factor estimated using our approach. The estimates are compared with the mean luminance of the HDR images scaled by a factor estimated from measurements.

Further, the assumption about mean luminance can be easily broken when the spot/multi-zone metering modes are used. Let's consider typical candidate scenes for HDR outdoor images: portrait or trunks of trees on the background of sky. It may happen that the photographer will be interested in the foreground or background. This could lead to significantly different automatic exposure settings (Figure 5.7). Other sources of errors for exposure-based approaches are discussed in [Wüller and Gabele 2007].

We compare our approach with the exposure based approach (5.18) as well as Ackermann et al. [Ackermann and Goesele 2013] modification of this equation. They suggest to account for the gamma curve of the sRGB colour space. The radiance value L_p for each pixel p with intensity value I_p can be found from:

$$L_p = 15.4 \frac{f^{-1}(I_p/255)}{f^{-1}(118/255)} \frac{N^2}{St}, \tag{5.19}$$

where f^{-1} corresponds to the inverse gamma correction of sRGB colour space. This approach can not be used with HDR images, so we estimate luminance values for each pixel for the middle image of the HDR sequence, which in our case corresponds to an automatic exposure setting. From the

absolute scaled values we then find the mean luminance of a scene. In order to get a scale factor for the HDR image, based on this approach, we divide the estimated absolute mean luminance by the mean luminance of the HDR scene.

We evaluate all the approaches on our clear sky dataset, where the image with the highest dynamic range has a range of 10 stops. All the images were captured using average metering mode. The images do not contain close up objects and are similar to the one in Figure 5.8c.



a. Clear sky, HDR Photographic Survey

b. Overcast sky, HDR Photographic Survey

c. Clear sky, our dataset

Figure 5.8: Example images from our dataset and HDR Photographic survey, which are used in the validation.

The value of K in (5.18) varies typically between 10.6 and 15.5. We keep K equal to 15.4 in (5.19) as suggested by Ackermann et al. We set K to 12.5 in (5.18) as the optimal value for our camera. We use the standard output sensitivity definition in both EXIF-data approaches as the one corresponding to our camera. Thus, our settings for equation (5.18) corresponds to the case of a fully known camera specification.

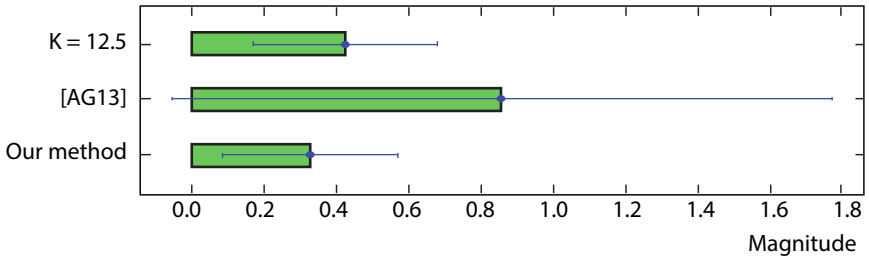


Figure 5.9: Mean and standard deviation of errors on images with clear sky regions from our dataset. Our approach is compared with [Ackermann and Goesele 2013] (AG13) and (5.18).

The images in our dataset meet the requirements imposed by EXIF-only methods. Under such conditions, EXIF methods yield correct absolute luminance values. The comparison of mean errors for our clear sky images is given in Figure 5.9, showing that our approach is statistically equivalent to the EXIF approach. This means that we deem our results to be accurate as well.

However, EXIF-only approaches are susceptible to error, for instance when other types of light metering are used. In the example of Figure 5.7 and Table 5.7 we show that the assumption about mean luminance can be easily broken for scenes with high dynamic range and depending on photographer intention. The same table shows that our method is robust against such variations, and therefore offers a significant advantage over EXIF-based methods.

5.4.2 Overcast Sky Images

Due to the limited number of overcast skies in either the HDR Photographic Survey and our dataset, we pool the results from both image collections.

A comparison of relative errors for scale factors estimated using approaches II.a, b or c is shown in Figure 5.10. Approach II.a gives the lowest overall error. Moreover, this approach requires fewer computations than II.b and c.

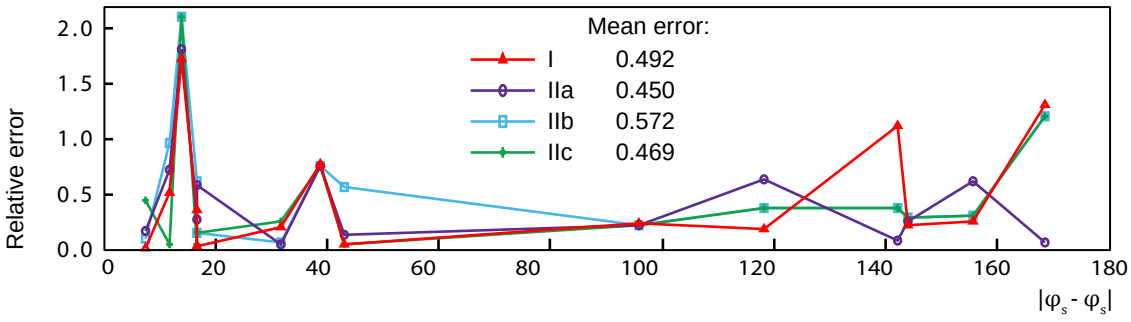


Figure 5.10: Relative errors for images with overcast skies as function of the azimuth difference between sun and camera. Here we compare various algorithms that estimate the type of overcast sky. For step I, we used $l_z \leftarrow l_{z_{\text{est}}}$ in conjunction with an estimate for L_z obtained using step IIa.

This comparison shows that solving the optimization problem for a second time with sky parameters fixed to the parameters of the estimated sky type, Approach II.a, produces more accurate results, than using only step I.

5.4.3 Validation of Scale Factors for Angular Maps

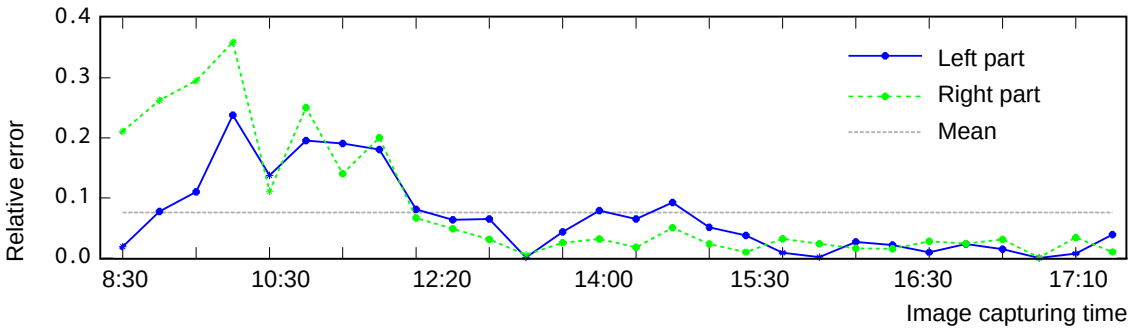


Figure 5.11: Magnitudes of relative errors of scale factors for the left and right parts of a sky probe to scale factors estimated from the whole sky probe.

To validate our algorithm on environment maps we used the sky probes from Stumpfel et al. [2004]. The absolute luminance values corresponding to the intensity values in the sky probes are unknown, so the images were divided into two equal parts. The scale factor was estimated for a whole image, as well as its left and right parts. Then, the relative errors of the scale factors from left and right parts to the scale factor from the whole image were evaluated (Figure 5.12). The mean error is less then ten percent for the full set of images (Figure 5.11). Results are in agreement with results of the validation of the algorithm on test images from our and Fairchild's datasets.

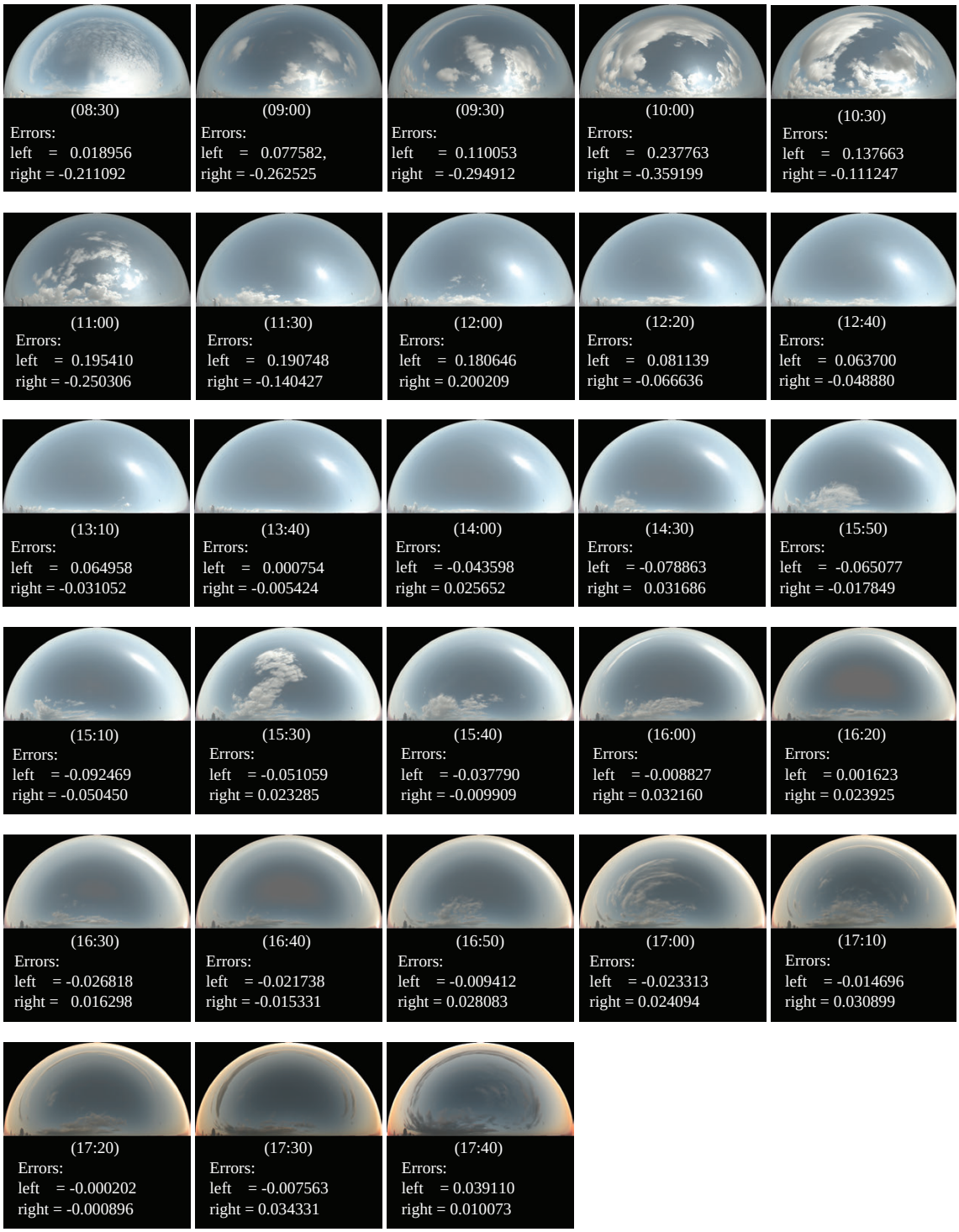


Figure 5.12: Relative errors for each of sky probe

5.5 The Hosek and Wilkie Sky Model

Recently, an extension for the Perez and Preetham sky models was suggested by Hosek and Wilki [2012]. This new model gives an improvement over the Perez sky model in modelling the sky appearance for lower solar elevation angles, as well as the appearance of the circumsolar ring. This is achieved by including two additional terms.

Moreover, the authors propose a new correspondence between turbidity values and parameters of the sky model, extending the range of its values from $[2,6]$ [Preetham et al. 1999] to $[1,10]$.

The introduced modifications require a zenith luminance model that is able to model the circumsolar ring. This is the reason why the authors, instead of normalizing by zenith luminance, find for each pair of turbidity and albedo values one “master value", which depends on sun elevation angle. In this form, unfortunately, the model can not be applied to our algorithm. Instead, we require the normalization by zenith luminance as in (5.10), with the distribution of luminance values that two additional terms provide.

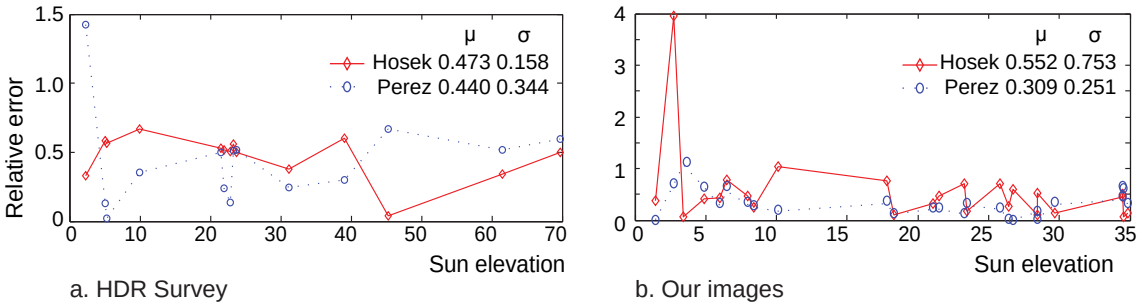


Figure 5.13: Comparison of magnitudes of relative errors for images with clear sky regions from Fairchild’s dataset [Fairchild 2007] in cases of Perez et al. [1993] and Hosek and Wilkie [2012] sky models using weighted combination of zenith models from [Karayel et al. 1984] and [Soler and Gopinathan 2000]. More comparison results can be found in the supplemental.

The results show that for our algorithm this model does not give advantages over using the Perez sky model (Figure 5.13). For images from our dataset, the Hosek model performs slightly worse than the Perez sky model. For images from Fairchild’s dataset, the Hosek model does not give any advantage in terms of mean of errors, but obtained results have a very low deviation from the mean. Given the increased number of required parameters, combined with the minimal improvements in our context, we have opted to retain the Perez sky model for our solution. The lack of improvement of the scale factor estimation using a new model is explained by the fact that we don’t rely on the regions where the improvement of sky representation was achieved. The region close to horizon is quite often occluded on the images. The region near the sun is automatically excluded on the sky segmentation step since without special equipment the SLR camera can not capture this region properly.

5.6 Discussion and Limitations

We validated our algorithm on images with diverse sky conditions and covering various sky areas (Figure 5.8). A limitation arises when the angle covered by the visible sky region is too small to contain sufficient information about sky intensity gradation. This is especially the case when the

cloud appearance is highly complex.

It is often thought that veiling glare could have an adverse effect on radiometric measurements of HDR images. In our specific case, however, we analyse skies, which often exhibit a fairly uniform or low frequency luminance distribution. In addition, we typically measure high luminance values that are integrated over large image areas. As a result, we deem the influence of veiling glare on our measurements to be marginal.

Finally, the accuracy of scale factor estimation will increase if time-lapse sequences are used [Lalonde et al. 2010]. Moreover, it could allow us to remove the requirement of known focal length and time of image capturing (information currently obtained from EXIF data).

Chapter 6

Gloss Editing in Light Fields

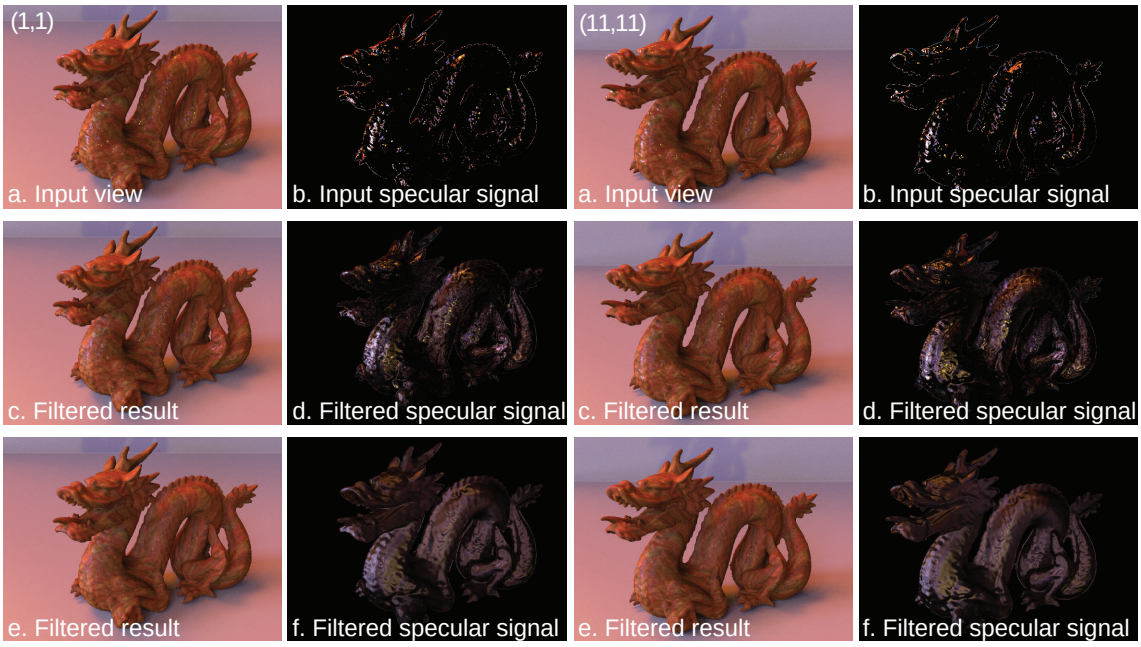


Figure 6.1: Our multidimensional filtering approach allows angularly-coherent editing of glossy appearance in a light field by means of a user-specified roughness parameter σ_v. *In pairs, from top to bottom:* input light field, and two editing results showing increasing roughness ($\sigma_v = 0.016$ and 0.064, respectively). Each pair shows the actual light field view, and the corresponding specular component with increased contrast for better visualization. We show here two views of the light field (left and right pairs of images).

Recent work on light field acquisition [Ng et al. 2005, Venkataraman et al. 2013], compression [Vetro et al. 2011], and display [Wetzstein et al. 2012], makes a good prospect for developing an end-to-end light field pipeline in the near future. An important component of such a pipeline is a post-processing stage where artists can edit the content to alter its appearance. In this chapter, we address the problem of material editing, and more precisely, editing the apparent roughness of a surface, in structured light fields.

In this context, several techniques have been proposed for editing of 2D content [Khan et al. 2006, Boyadzhiev et al. 2015], but the problem remains mostly unexplored for light fields (Section 3.3).

Although it seems feasible to use existing 2D techniques and apply them separately to different views, such an approach does not exploit the rich information encoded in light fields. As a result, complex interactions between scene geometry, viewing location, and light positions cannot be easily handled. As discussed in Section 3.2.1, in dynamic conditions the gloss perception is stronger than for static images. Thus, the editing should be performed consistently with surface geometry and between views. For instance, when the highlight size is increased, this highlight might get visible in a different view, that requires propagation of edits between neighbouring views.

Even though light fields enable full scene reconstruction by means of inverse rendering [Haber et al. 2009], the problem is usually difficult, and requires solving a highly underconstrained optimization problem. Fortunately, the accurate solution might not be needed. Since reflectance estimation is an underconstrained problem, the human visual system (HVS) does not perform inverse rendering to obtain information about physical properties of materials (Section 3.2). That suggests that an approximate solution for material editing might be sufficient.

Inspired by these findings, our main observation is that an increase in surface roughness leads to changes in the size and brightness of specular highlights, and that such operations can be performed using simple image processing techniques, and in particular filtering. Consequently, we propose a new light field editing approach which modifies the apparent roughness of the surface by filtering the specular component of the light field. Since angular consistency is of high importance for material perception (Section 3.2.1), filtering is performed jointly in the spatial and angular domains, with parameters optimized to closely match ground-truth results. Inspired by perceptual findings, our approach explicitly avoids the difficult problem of light source position estimation, and relies only on the depth map, which can be computed for light fields with a satisfactory fidelity (Section 2.6). For this purpose, we adapt existing solutions paying special attention to the depth reconstruction in highlights, which are often prone to artefacts (Section 6.1). We demonstrate that even partial scene information, such as depth and surface normals, enables high quality editing of the object appearance (Section 6.2). One of the key benefits of using a filtering approach is that in difficult regions it provides graceful quality degradation without catastrophic failures. As a result, the edits do not introduce visible artefacts and inconsistencies (Section 6.3).

6.1 Light Field Preprocessing

To run our gloss editing algorithm we require normal maps of the light field views, as well as a separation of the specular and diffuse components. In this section we describe how depth maps can be estimated, how normals can be computed from these depth maps, and how to obtain the separation for the specular and diffuse components.

6.1.1 Depth Estimation

There has been a significant progress in methods estimating depth information from light field data (Section 2.6).

We use the approach by Chen et al. [2014] as it provides good reconstructions at object edges, which is important to our algorithm. Moreover, in a recent work [Williem and Kyu Park 2016] it was shown to perform the best in absence of noise. After computing depth map, we perform an additional step that improves normal estimation and angular coherence. For completeness, we provide details of our implementation of the method by Chen et al., followed by details of our

Symbol	Meaning
d	The disparity value
$\mathbf{d}$	The disparity maps, containing disparity values for each point of the input light field
$S \equiv S_{s,t,x,y,d}$	The SCam computed from the input light field for the scene point observed in the input light-field point with coordinates (s,t,x,y), assuming the disparity value equal to d (Equation 2.26)
$S^*_{s,t,x,y,d}$	The instances of the scene point within $S_{s,t,x,y,d}$ viewed in the input light-field point with coordinates (s,t,x,y)
$\rho_d(s,t,x,y)$	The photo consistency measure computed on the corresponding $S^*_{s,t,x,y,d}$
$\mathbf{c}$	The global confidence measure of disparity estimates
$S^d_{s,t,x,y,d}$	The SCam computed from the disparity maps $\mathbf{d}$ for the scene point observed in the input light-field point with coordinates (s,t,x,y), assuming the disparity value equal to $\mathbf{d}(s,t,x,y)$ (Equation 6.3)
$S^c_{s,t,x,y,d}$	The SCam computed from the consistency maps $\mathbf{c}$ for the scene point observed in the input light-field point with coordinates (s,t,x,y), assuming the disparity value equal to $\mathbf{d}(s,t,x,y)$ (Equation 6.4)

Table 6.1: Symbols used in the description of the algorithm.

extension.

Disparities Estimation

Approach of Chen et al. [2014] belongs to the group of approaches that compute photo consistency on SCam images (Section 2.6.2).

As discussed in Section 2.6.2, Chen et al. [2014] introduced a bilateral consistency measure (BCM) to identify the set of pixels $S^*_{s,t,x,y,d}$ within the SCam $S_{s,t,x,y,d}$ which represents the same scene point:

$$BCM_{s,t,x,y,d}(s_i,t_i) = e^{-\frac{\Delta_c}{2\sigma_c^2} - \frac{\Delta_s}{2\sigma_s^2}}, \tag{6.1}$$

where $\Delta_c = \left(S_{s,t,x,y,d}(s_i,t_i) - S_{s,t,x,y,d}(s,t)\right)^2$ and $\Delta_s = ((s_i,t_i)-(s,t))^2$, (s,t) is the reference point. We set the colour variance σ_c^2 and the spatial variance σ_v^2 to $5/255$ and $\lfloor \max(N_s,N_t)/2 \rfloor$, correspondingly. The final sought-for set S^* is obtained by thresholding values of BCM. Afterwards, the final photo-consistency measure is computed on the set S^* for each trial disparity value $\tilde{d}$:

$$\rho_{\tilde{d}}(s,t,x,y) = \frac{1}{|S^*_{s,t,x,y,\tilde{d}}|} \sum_{(s_i,t_i)\in S^*_{s,t,x,y,\tilde{d}}} e^{-\frac{\Delta_c}{2\sigma_c^2}}, \tag{6.2}$$

where $|S^*_{s,t,x,y,\tilde{d}}|$ is the size of the set $S^*_{s,t,x,y,\tilde{d}}$. The initial disparity values are obtained as the optimal value of the consistency measure: $\mathbf{d}(s,t,x,y) = \underset{\tilde{d}}{\operatorname{argmax}}\, \rho_{\tilde{d}}(s,t,x,y)$. Following Chen et al. [2014] we as well account for the local spatial coherence. Please see [Chen et al. 2014] for the details.

This initial depth estimation computes disparity estimates $\mathbf{d}$ for each point of the input light field. The estimation might be noisy and angularly inconsistent. It may also contain errors in

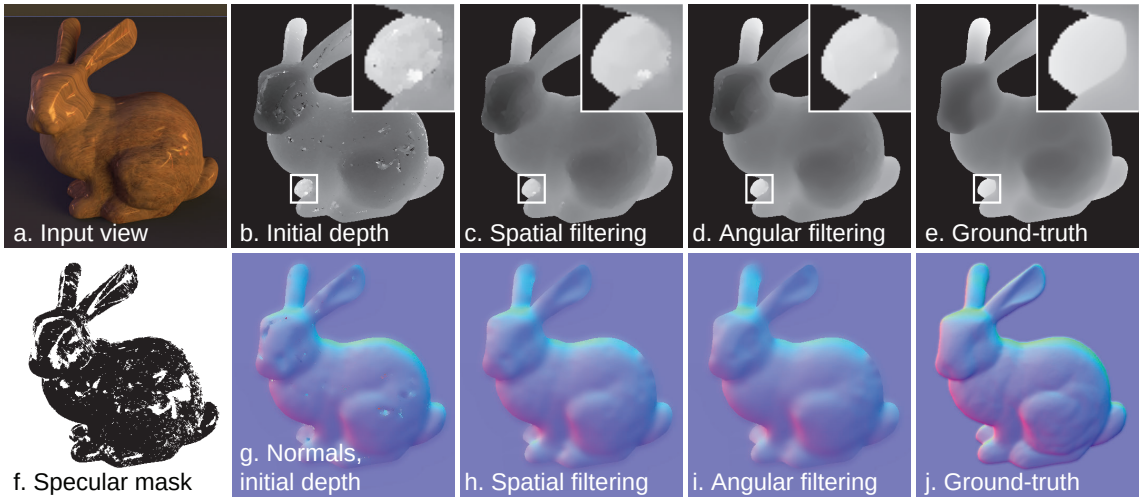


Figure 6.2: Illustration of the depth maps for each of the steps in Section 6.1.1. The input light field consists of 15x15 views. The second row shows our estimated mask of the specular regions and the normal maps, computed as in Section 6.1.2. The maps show the values in the masked region, but the masks are not used for the steps in Section 6.1.

specular regions (see Figure 6.2b). Most solutions which perform regularization or filtering to avoid these problems involve a MRF formulation to solve an energy minimization problem [Tao et al. 2015b, Wang et al. 2016, Lin et al. 2015]. While global regularization can further improve spatial consistency of depth estimates and remove noise, we exploit a computationally less expensive approach. Following Chen et al. [2014], we identify pixels with low confidence of their depth estimates and resort to spatial depth propagation in these regions. The global confidence $\mathbf{c}$ is computed as the ratio of two quantities: the difference between the global and the second biggest local maximum of the photo-consistency measure c and the difference between the global and the smallest local maximum. The *unreliable* pixels can be defined as the ones with $\mathbf{c} < \theta$, where θ is a threshold value (in our implementation $\theta \in [0.8, 0.9]$). Once we have detected the regions where depth is inaccurate, we use a strategy similar to [Lin et al. 2002] to propagate depth from reliable to unreliable regions within each view of the light field. For this propagation we rely on the idea that the depth of unreliable pixels can be estimated as the average of depth values of reliable pixels in some window around the unreliable pixel. We achieve this with a push-pull algorithm. To this end, we first compute a Gaussian pyramid based on reliable pixels, and then, propagate depth information in a top-down fashion to unreliable pixels from coarser levels. As a result, the depth map is significantly improved (compare Figures 6.2b and 6.2c). This type of propagation can lead to errors on objects edges. Moreover, within-view propagation might lead to angularly inconsistent depth estimates. However, most regularization or filtering solutions are designed to work *within each view separately*, since most depth estimation algorithms seek a depth map only for the central view.

To obtain angular consistency and alleviate the errors on the edges, we perform an angular filtering step, which is performed *within SCams*. We exploit the fact that the scene points which have a specular reflection in certain views will not have them in some others. For points with specular reflection, disparity estimates are recomputed through within-view propagation. The estimates obtained in this way are less reliable than the one computed with high confidence as optimal values of photo-consistency measure. Therefore, our goal is to propagate angularly the disparity estimates with high values of the consistency measure to those, where the disparity has been estimated via

within-view propagation.



Figure 6.3: SCams computed from the input light field, depth and confidence maps. **(a)** An SCam computed for the scene point corresponding to the light-field point with coordinates $(t,s,y,x) = (2,11,145,250)$ (the point is indicated with a green star). The view $(t,s) = (2.11)$ is shown in Figure 2.7a. The SCam is computed with a ground-truth depth value. **(b)** An SCam computed for the same scene point as (a) from the initial depth maps (Equation 6.3). This SCam demonstrates the consistence of depth estimates between different views. The image is normalised to a range from 0 to 1 for a better visualization. **(c)** An SCam computed from the confidence maps for the same scene point as (a) (Equation 6.4). **(d)** An SCam computed from the final depth maps for the same scene point as (a). The computation is performed as in Equation 6.3, where instead of the initial disparity maps the ones obtained after the angular propagation are used. The image is normalised to a range from 0 to 1 for a better visualization.

For each reliable pixel (x,y) within each view (s,t), we compute the SCams from the disparity $\mathbf{d}$ and the confidence $\mathbf{c}$ maps (Figure 6.3b,c):

$$S^d \equiv S^d_{s,t,x,y,d} = \bigcup_{(\Delta s,\Delta t)} \mathbf{d}(s+\Delta s, t+\Delta t, x+\Delta s d, y+\Delta t d) \tag{6.3}$$

$$S^c \equiv S^c_{s,t,x,y,d} = \bigcup_{(\Delta s,\Delta t)} \mathbf{c}(s+\Delta s, t+\Delta t, x+\Delta s d, y+\Delta t d), \tag{6.4}$$

where $d = \mathbf{d}(s,t,x,y)$. Instead of propagating d, we first compute an average $\bar{d}$ in S^d over the reliable pixels with similar disparity values: $\bar{d} = (\sum_j w_{ij} S^d(j))/\sum_j w_{ij}$, where i is the index in S^d of the pixel (s,t,x,y). This step is performed to account for small errors in disparities estimation obtained as optimal values of the consistency measure ρ. During the averaging, the pixels are weighted by:

$$w_{ij} = e^{-\frac{w_d^2}{2\sigma_d^2}} \max\left(0, \frac{S^c(j)-\theta}{1-\theta}\right) \max\left(0, \frac{S^c(i)-\theta}{1-\theta}\right), \tag{6.5}$$

where w_d is the difference between disparity estimates, σ_d defines how similar two disparity values should be to be considered as two estimates for the same scene point. The other terms ensure we do not take into account pixels with unreliable depth estimates (w_{ij} is zero if confidence value $S^c(j)$ or $S^c(i)$ is less than the threshold θ).

Once the averaged disparity value $\bar{d}$ is obtained we want to propagate it to the other pixels $j \in S^d$:

$$S^d(j) = \bar{d} w^p_{ij} + S^d(j)(1-w^p_{ij}), \tag{6.6}$$

where we perform linear interpolation between the average $\bar{d}$ and the old disparity value $S^d(j)$ based on how reliable and how similar its value to $S^d(i)$. To do so, we now compute the weight w^p_{ij} for propagation:

$$w^p_{ij} = e^{-\frac{w_d^c}{2\sigma_d^2}\frac{\theta^2}{(2S^d(j))^2}} \max\left(0, \frac{S^c(i)-\theta}{1-\theta}\right), \tag{6.7}$$

where the exponential term determines to which pixel to propagate, and the second term as before states that the propagation should not be done if $i = (s,t,x,y)$ has low value of confidence measure. The term $\frac{\theta^2}{(2S^d(j))^2}$ influences the value of σ_d, so that the disparity values of unreliable pixels j could differ from $S^d(i)$ more.

6.1.2 Normals Computation

The technique described in Section 6.1.1 computes a disparity value for each pixel of each view in the light field, $\mathbf{d}$. To compute the corresponding normals, we first convert the disparity value $d \in \mathbf{d}$ to depth value Z using Equation (2.24). We further compute horizontal (Gx) and vertical (Gy) gradients using forward differences. The tangent plane for a point is given by vectors $[1,0,Gx(s,t,x,y)]$ and $[0,1,Gy(s,t,x,y)]$, and the normal is computed as the cross product of these two vectors. Finally, we apply a guided filter [He et al. 2010] to the computed normal map with the depth map as a guidance image, to obtain a smooth normal map, and normalize vector lengths to one (Figure 6.2 g-j).

6.1.3 Specular and Diffuse Component Separation

For a review of different approaches for specular separation please refer to [Artusi et al. 2011]. We select an approach which does not require any assumptions on illumination or any type of hallucination for missing information [Criminisi et al. 2005, Tao et al. 2015a].

The specular separation is performed under the assumption that the point containing the specular reflection in one view will be observed as diffuse in some other. We further assume a Lambertian model for the diffuse component of the surface BRDF. To find all the instances of the point (s,t,x,y) with disparity value $d = \mathbf{d}(s,t,x,y)$ one needs to find the region $S^* \equiv S^*_{s,t,x,y,d}$ (Section 2.6.2). Compared to a task of an initial segmentation of an SCam for the purpose of depth estimation (Section 2.6.2) we now know the disparity estimates $S^d_{s,t,x,y,d}$ (Equation 6.3) and can immediately obtain S^* by taking the pixels with depth estimates similar to d (Figure 6.3). To an accuracy of noise level, the diffuse component can be estimated as the RGB value of the pixel with the lowest intensity value within S^*. Following Lin et al. [Lin et al. 2002], we consider all unreliable pixels, defined in Section 6.1.1, as specular.

6.2 Filtering

We pose the process of editing glossy appearance as a filtering problem, in which the edited value of each pixel in the light field is computed from the values in neighboring pixels according to a *roughness* parameter specified by the user. In a light field, neighboring pixels refer to both angular and spatial neighbors, i.e., filtering should be done in both the spatial and angular dimensions, combined, since the shape and appearance of the highlight is affected by both domains. We describe and justify the filtering approach below. Note that we filter only the specular component of the input light field (or of a masked region of it). The object in light field can be selected for example using an interactive technique [Wanner et al. 2013]. Thus, when we refer to light field values we will be referring only to the specular component of the light field. However, we maintain the notation of the previous section for clarity.

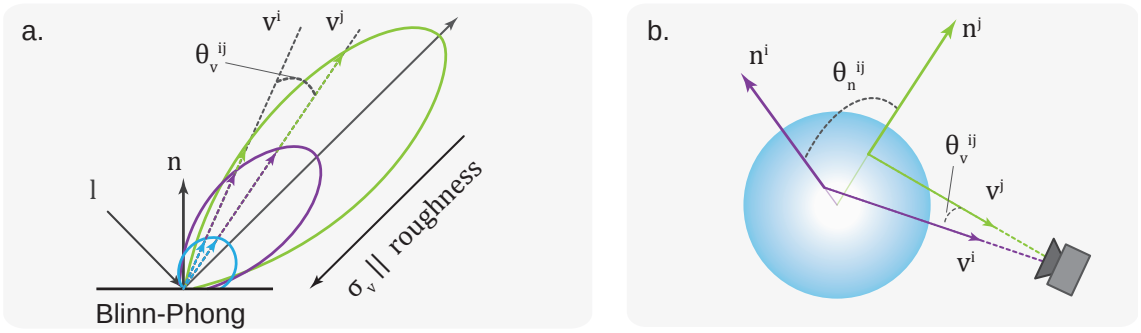


Figure 6.4: **(a)** Reflectance lobes at a certain scene point, with normal n and light direction l, for different roughness values. Vectors v^i and v^j illustrate view directions of the scene point to two views of the light field, i and j. **(b)** Normals and view directions for two scene points within one view.

Angular filtering We exploit here that the light field contains instances of a scene point as observed from different view directions. For each point (s,t,x,y) of the input light field L with the disparity $d = \mathbf{d}(s,t,x,y)$, these instances form the set $S^* \equiv S^*_{s,t,x,y,d}$, defined in Section 2.6.2. The pixels of S^* share the same depth, normal, and directions to the light sources, while the direction to each camera (or viewpoint) v_j varies.

Our task is to obtain the new intensity value for a point from the input light field, which implicitly encodes information about the light sources, and the known information about varying view directions.

This can be formulated as a convolution of the input specular signal with some view-dependent weights, yielding the edited signal $\tilde{S}^*$. This convolution is done *per SCam of the light field* as follows:

$$\tilde{S}^*(i) = \sum_{j \in \hat{S}^*} \hat{S}^*(j) k^a_{ij}(\sigma_v), \quad \text{for each } i \in S^*, \tag{6.8}$$

where $\mathbf{k}^a$ is the convolution kernel, $\hat{S}^*$ are the pixels of S^* with values greater than zero. For the computation of the kernel weights k^a_{ij} we consider the following:

(i) *two pixels i and j should share more information (i.e., weights will have higher values) if their view directions are more similar,*

(ii) *the higher the roughness value we want to achieve, the more information will be shared between the pixels with similar view directions.*

This is illustrated in Figure 6.4a, where we depict reflectance lobes for different roughness values, together with view directions for a certain scene point as captured in two views of the light field, v^i and v^j. For the different roughness values, one can see that, the smaller the angle $\theta_v^{i,j}$ between camera directions ($\theta_v^{i,j} = \arccos(v^i \cdot v^j)$), the more similar the reflectance values of the pixels will be; and that, the higher the roughness, the more similar the reflectance values as well, as explained in (i) and (ii) above. Taking this into account, the weights of the convolution kernel for the computation of pixel i are defined as:

$$k^a_{ij}(\sigma_v) \equiv k^a(v_i, v_j, \sigma_v) = \exp\left(-\frac{\theta_v^{i,j^2}}{2\sigma_v^2}\right) \omega_{ij}, \tag{6.9}$$

where σ_v relates to the desired roughness. For the base version of our weights we set ω_{ij} to 1; its actual value is introduced next.

The question arises if there is any other information which can be used without implicit computation of the scene lighting. Let us consider the widely used Cook-Torrance BRDF model:

$$\rho(i,l,v) = \frac{D(l,v)F(l,v)G(l,v)}{4(v\cdot n)(l\cdot n)}, \tag{6.10}$$

where $D(\cdot)$ is a microfacets distribution function, $F(\cdot)$ is a Fresnel term, $G(\cdot)$ is a geometrical attenuation term, l is the light source direction, and i is some scene point. All the terms, except for the $(v\cdot n)$ term in the denominator, depend on the light source direction. While performing a convolution we can account for this term by multiplying the input signal at j by $(v^j\cdot n^j)$ and dividing the result of the convolution by $(v^i\cdot n^i)$. That corresponds to setting weights ω_{ij} in Equation (6.9) to the following:

$$\omega_{ij} \equiv \omega(v_i,v_j) = \frac{(v^j\cdot n^j)}{(v^i\cdot n^i)}. \tag{6.11}$$

Spatial filtering For different points within each view of the light field, the normals, camera directions and light source directions might differ. However, in a small neighborhood around a given surface point one can assume a constant light source direction. Thus, the similarity between pixels is determined not only as a difference between view directions, but also as a difference between normal vectors. This is also modelled as a convolution, involving view- and normal-dependent kernel weights k^s_{ij} and yielding the edited signal $\tilde{L}_{s,t}$. This convolution is done *per view of the light field* as follows:

$$\tilde{L}_{s,t}(i) = \sum_{j\in\hat{L}_{s,t}} \hat{L}_{s,t}(j)k^s_{ij}(\sigma_v,\sigma_n), \quad \text{for each } i\in L_{s,t}, \tag{6.12}$$

where $\hat{L}_{s,t}$ are pixels of the view $L_{s,t}$ with values greater than zero. The kernel weights are computed as follows:

$$\begin{aligned} k^s_{ij}(\sigma_v,\sigma_n) &\equiv k^s(v_i,v_j,n_i,n_j,\sigma_v,\sigma_n) = \\ &= \exp\left(-\frac{(\theta^{i,j}_v)^2}{2(\sigma_v)^2} - \frac{(\theta^{i,j}_n)^2}{2(\sigma_n)^2}\right)\omega_{ij}, \end{aligned} \tag{6.13}$$

where $\theta^{i,j}_n = \arccos(n_j\cdot n_i)$, $\theta^{i,j}_v$ is the same as in Equation(6.9), but computed over a different set of light field points: $i\in L_{s,t}$ and $j\in\hat{L}_{s,t}$ (see Figure 6.4b). Parameters σ_v and σ_n control how rough the final material will appear in the light field: *the higher the roughness value we want to achieve, the more information will need to be shared between the pixels with similar view directions and normal vectors*. This is achieved by increasing σ_v and σ_n.

Combined Filtering One can observe that in the case in which i in Equation (6.13) is set to go over S^*, and j to go over $\hat{S}^*$, then $k^s(\sigma_v,\sigma_n) = k^a(\sigma_v)$, since in this case $\theta^{i,j}_n = 0$. This allows us to define the universal kernel weights $k_{ij}(\sigma_v,\sigma_n)$ by Equation (6.13), such that

- $k_{ij}(\sigma_v,\sigma_n) = k^a(\sigma_v)$ if $i\in S^*$ and $j\in\hat{S}^*$,
- $k_{ij}(\sigma_v,\sigma_n) = k^s_{ij}(\sigma_v,\sigma_n)$ if $i\in L_{s,t}$ and $j\in\hat{L}_{s,t}$.

In this case the convolution is done *per LF* and takes the form: $\tilde{L}(i) = \sum_{j \in \hat{L}} \hat{L}(j) k_{ij}(\sigma_v, \sigma_n)$, for each $i \in L$, where $\hat{L}$ are pixels of L with values greater than zero, and i denotes each of the pixels in light field L.

6.2.1 Filtering Parameters and Validation

In order to validate the multidimensional filtering, as well as to study the relation between σ_n and σ_v, we search for optimal values of these parameters by formulating an optimization problem.

Optimal filtering parameters We minimize the luminance differences between a target material with roughness r_2, and an input material with roughness r_1 filtered to obtain the target one. To avoid convergence to local minima, this optimization is performed in two steps: We first solve for σ_v and then for σ_n.

When solving for σ_v, we only perform *angular* filtering. Further, we select SCams $\mathcal{S}$ that contain the peak of the specular reflection. This ensures that our optimization results will not be biased by the SCams with partial information about lighting. We demonstrate in the next section that such SCams result in missing energy in the filtered light field. Consequently, we minimize the luminance differences between the specular component of the target SCam, $S^*_{r_2}$, and that of the filtered SCam, $\tilde{S}^*_{r_1}$, for all SCams in subset $\mathcal{S}$:

$$\min_{\epsilon, \sigma_v} \sum_{S^*_{r_2} \in \mathcal{S}} \sum_{i \in \tilde{S}^*_{r_2}} \left(S^*_{r_2}(i) - \epsilon \sum_{j \in \hat{S}^{r_1}} k^a_{ij}(\sigma_v) \hat{S}^*_{r_1}(j) \right)^2, \tag{6.14}$$

where ϵ is a global normalization factor. Once we have σ_v and ϵ, we obtain σ_n by minimizing the differences over the specular regions of the full light field:

$$\min_{\sigma_n} \sum_{i \in L^*_{r_2}} \left(L_{r_2}(i) - \epsilon \sum_{j \in \hat{L}_{r_1}} \hat{L}_{r_1}(j) k_{ij}(\sigma_n, \sigma_v) \right)^2, \tag{6.15}$$

where k_{ij} is given by Equation (6.13). We solve Equations (6.14,6.15) using the implementation of the trust region-reflective algorithm from the MATLAB Optimization Toolbox.

We solved these optimization problems for several target roughness values to uniformly cover the range of possible roughness values and observed that consistently we obtained $\sigma_n \approx \sigma_v/2$. That is in accordance with the following intuition: If we consider the case where the normal, light source direction and view direction are lying on the same plane, a change of the normal in this plane by some angle γ will lead to the half angle changing by γ. At the same time, a change of a view direction by γ leads to a change in the half angle by $\gamma/2$. To obtain all results in Section 6.3, we fix the ratio between these parameters, so that $\sigma_n = \sigma_v/2$.

Validation To assess the contribution of our weighting factor ω_{ij} (defined in Equation 6.11), in Figure 6.5 we show the results of filtering for different target roughness values, with and without this factor. In the figure we depict a 1D SCam of the light field. Note that without this weighting factor the filtering can not achieve the asymmetry which is inherent to large roughness values.

Once we have shown that the weighting factor ω_{ij} is required to properly depict the appearance, we evaluate the contribution of the different types of filtering, i.e., considering only angular, spatial, or

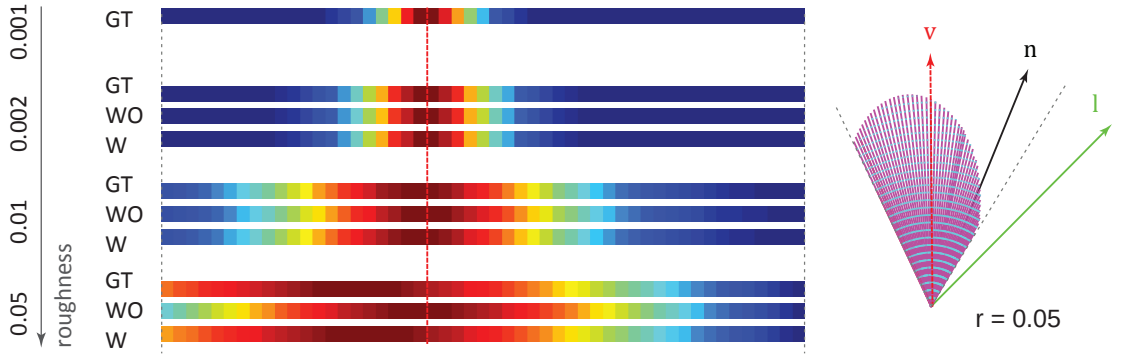


Figure 6.5: Evaluating the contribution of the weighting factor ω_{ij} on a 1D SCam of a light field with $N_s = 50$. **Left:** An input light field with a roughness $r = 0.001$, is filtered with different parameters to obtain different target roughnesses (0.002, 0.01 and 0.05). For each target roughness, the three 1D SCams show the target (ground truth) SCam (GT), the filtered SCam without ω_{ij} (WO), and the filtered SCam with ω_{ij} (W). Note that ω_{ij} is required to capture the asymmetry of the ground truth SCam. **Right:** Illustration of light reflected at the points sampled by the 1D SCam (for the specific case of $r = 0.05$), for a scene point with normal n and light direction l (violet for actual values and cyan for filtered values).

combined filtering. In Figure 6.6 we show the results obtained with each type. If angular filtering

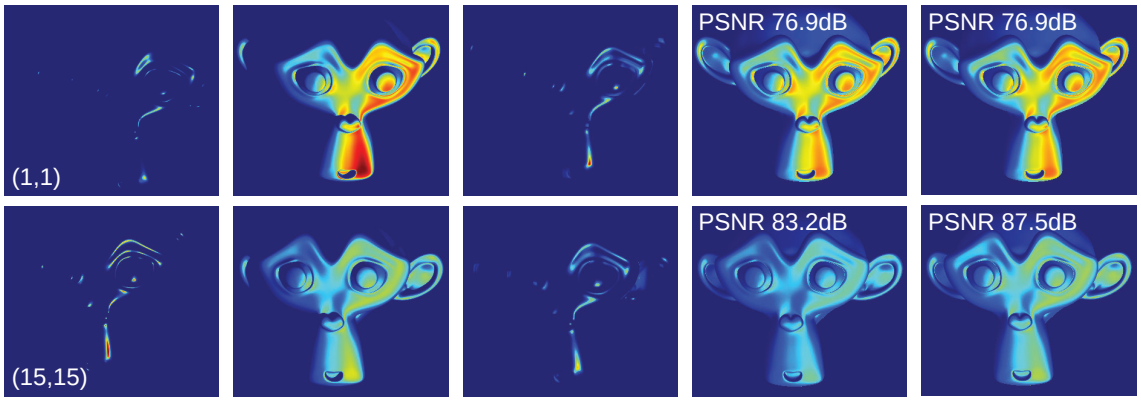


Figure 6.6: Evaluation of angular, spatial and combined filtering on a 4D light field (15x15 views; the figure shows the heatmaps for the luminance channel of the views (s,t)=(1,1)/(15,15)). The input light field **(A)** has a roughness value r = 0.001, and the target roughness is 0.1 **(B)**. Values of σ_v and σ_n are obtained as described in Section 6.2.1. The light fields **(B,D,E)** are normalised to match the mean luminance of the view (s,t) = (1,1).

alone is performed, for it to be correct, an SCam should cover the full specular lobe. Otherwise, filtering will lead to a lack of energy compared to the ground truth (see Figure 6.6C vs. B). Spatial filtering within a single image allows to correctly modify the appearance of the specular component within a certain neighborhood. However, the closer the light source is to the surface, the less reliable the spatial propagation becomes for distant points. Thus, the spatial filtering might not be sufficient and may lead to inconsistent appearance of angular-dependent effects (see Figure 6.6 columns D vs. B). Accounting for information both angularly and spatially ensures a more correct distribution of energy over the light field (Figure 6.6 column E).

6.3 Results and Limitations

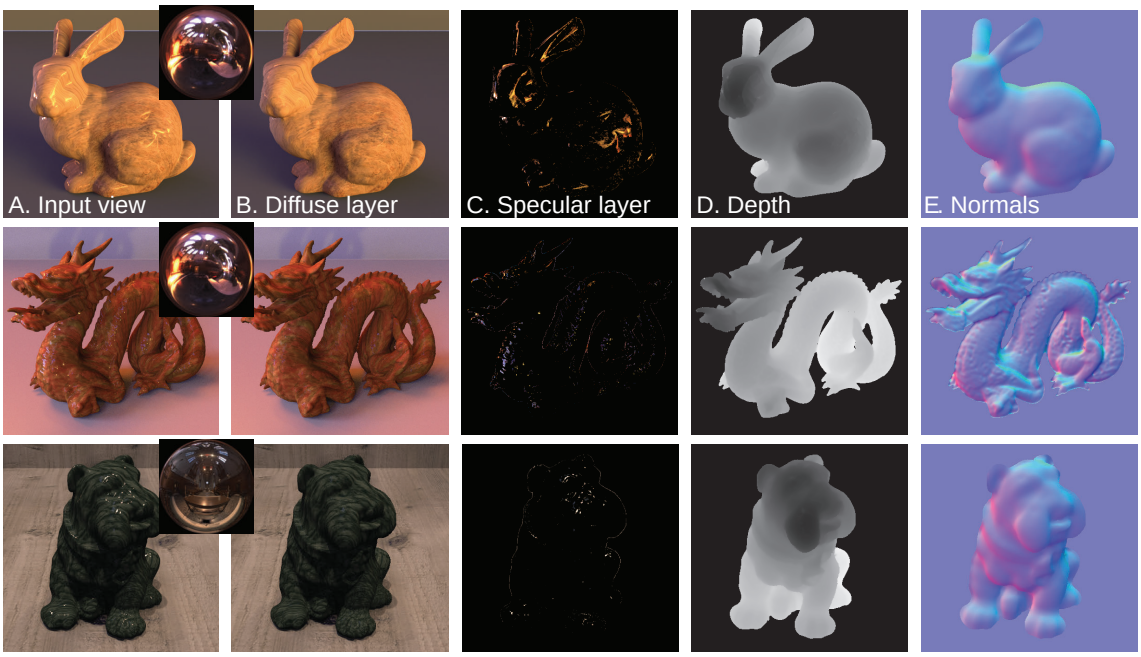
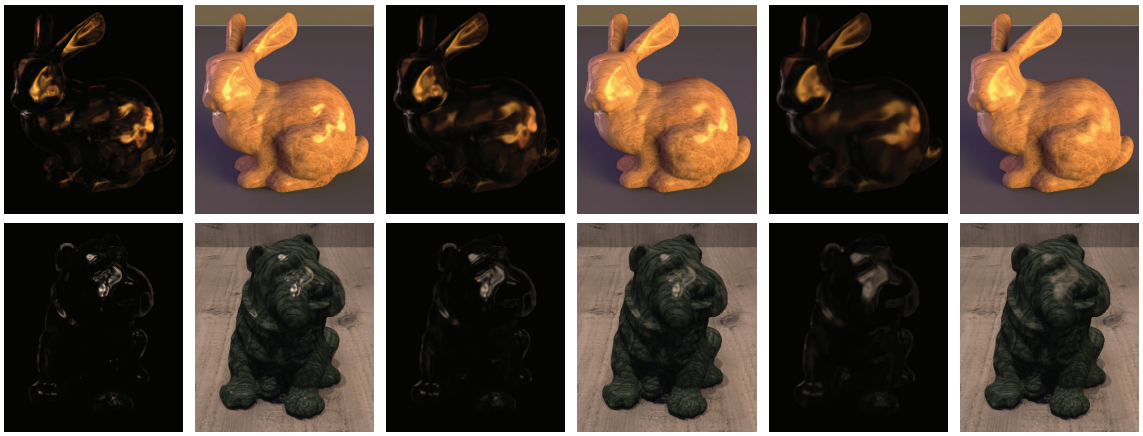


Figure 6.7: Steps of our preprocessing pipeline for three sample scenes: bunny, dragon and dog. All three input light fields have 11 by 11 views. For the bunny scene the disparity range is set to [-6:-3.3], for the dragon scene to [-6:-2.5] and for the dog scene to [-5.8:-2.5].

The full preprocessing pipeline is shown in Figure 6.7: Depth is estimated for each view (Figure 6.7E), followed by normals computation (Figure 6.7F) and separation of specular and diffuse components (Figure 6.7C and D). In Figures 6.8 and 6.1 we show filtering results for three different output roughness values for each of input scenes. For each result we show both the filtered specular layer and the composited result; when merging both layers, the filtered layer is scaled so that the mean values of the input and filtered specular layers match. Please refer to the supplemental material for videos showing the full light fields of final and intermediate steps for all scenes. Our technique yields angularly-coherent, plausible edits of the appearance of glossy objects. In Figure 6.9 and supplemental videos, we show side-by-side comparisons with ground-truth, as well as error maps computed using the SSIM [Wang et al. 2004] between the filtered and ground-truth results. Note that the major differences are caused by imprecision of depth reconstruction and not by the filtering.

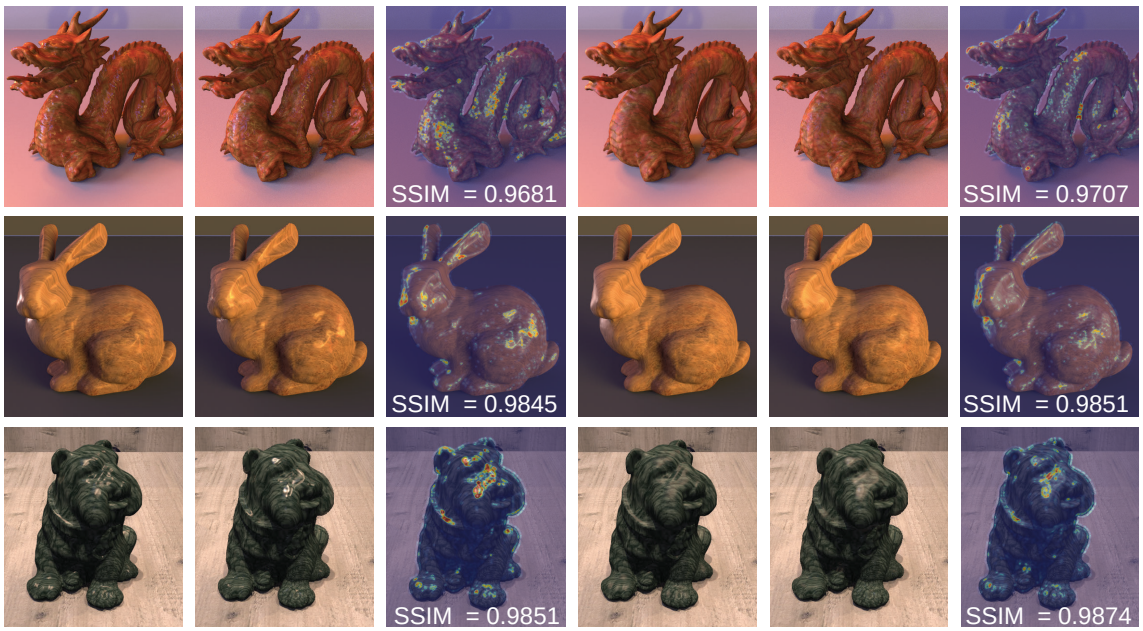
Additionally, varying the weight of the specular layer while compositing one can achieve more/less metallic appearance of objects (Figure 6.10 and supplemental).

We also show that simple 2D images manipulation is insufficient for editing a material's roughness. The recent work of Boyadzhiev et al. [Boyadzhiev et al. 2015] enables single-image modification of some material properties. In terms of gloss appearance it can make the material more or less shiny; we show results of both operations in Figure 6.11 (a and b) for our *dragon* scene. With this technique, not only the highlight is modified, but the texture is also affected (see, e.g., Figure 6.11b and supplemental). Further, while on a single image the results of both techniques may not be strikingly different, the difference becomes clear when observing the whole light field (please see supplemental videos).



A. Filtered: $\sigma_v = 0.016$ B. Filtered: $\sigma_v = 0.032$ C. Filtered: $\sigma_v = 0.064$

Figure 6.8: Several filtering results for the scenes in Figure 6.7, with varying degrees of output roughness. For each result we show both the filtered specular layer and the composited result.



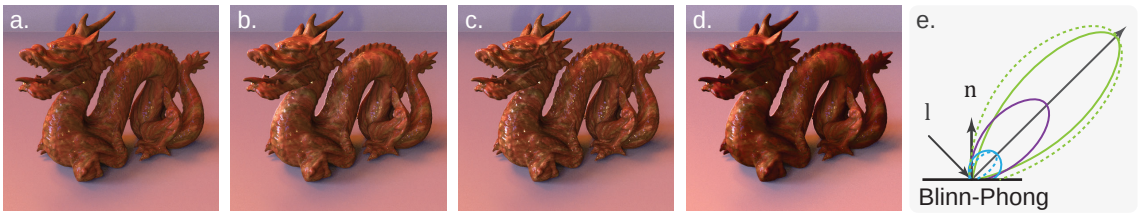
A. Ground-truth result | B. Filtered, σ_v=0.016 | C. Filtered + SSIM | D. Ground-truth result | E. Filtered, σ_v=0.064 | F. Filtered + SSIM

Figure 6.9: Comparison of the filtered results **(B,E)** with ground-truth results for different roughness values **(A,D)**. In this case no optimization of parameters was performed. The ground-truth results were selected manually by visual similarity to the filtered one. Figures **(C,F)** show the overlay of the filtering results ($\sigma_v = 0.016/\sigma_v = 0.064$) with the dissimilarity maps between the filtered and ground-truth results, computed using SSIM [Wang et al. 2004].



Figure 6.10: Edited results obtained merging the diffuse layer with a linearly scaled filtered specular layer ($\sigma_v = 0.032$).

Finally, in Figure 6.11c we plot the errors of this technique, under these assumption that the decomposition allows to process the highlight separately from the texture. The figure shows that when highlights get brighter or dimmer, the shape of the lobe does not change correctly. While our technique allows to model that behaviour correctly (Section 6.2.1 and Figure 6.5).



Input | Boost 10: ampl.: low, freq.:low, sign: pos. | Boost 2: ampl.: high, freq.:high, sign: pos. | Boost 5: ampl.: all, freq.:low, sign: pos.

Figure 6.11: **(a)** Input view. **(b,c,d)** Results of gloss editing using [Boyadzhiev et al. 2015]: the three operations, producing the results the closest to making the material more shiny, (the parameters used are specified in the figure, for their meaning please refer to [Boyadzhiev et al. 2015]). **(e)** Illustration of the scaling operations on an input reflectance lobe (purple), compared to ground truth specular lobes. Solid lines correspond to the ground truth, and dashed ones to the results obtained by scaling the lobe.

The major limitation of our approach comes from depth estimation and specular separation algorithms. Both, non-textured objects or objects with very large specular highlights, will result in poor depth estimation. Possible solutions involve integrating depth cameras, hallucinating missing information, or using additional hardware for, e.g., specular separation. So far our algorithm is targeted for roughness increase only and is applicable only to opaque objects. As mentioned in Section 6.2 the editing requires preliminary masking the object or the material. We use the ground-truth masks obtained during rendering. Alternatively, the masks can be obtained automatically, using segmentation techniques [Mihara et al. 2016].

Chapter 7

Conclusion

This chapter summarizes contributions of this thesis. First a general discussion is given, followed by a summary and contributions within each of the considered problems. We conclude with a broader discussion on future work stemming from reciprocity of HDR and light-field technologies in Section 7.2.

7.1 Summary and Outlooks

Despite significant progress in capturing and displaying technologies, a recreation of a real scene immersively remains a challenging problem.

Such aspects as temporal and spatial resolution play an important role. Tremendous progress was done in increasing pixel resolution. A frame rate of capturing devices is also increasing over time.

However, a much more important problem is the captured colour gamut, luminance range and contrast. Particularly, the captured colour gamut of an LDR image represents just a small part of the visible one. The conventional imagery's intensity range as well as contrast are far from matching the ones found in real world scenes. Capturing an HDR content with consumer-available cameras remains an open problem.

Further, while the world surrounding us is three-dimensional the conventional imaging is two dimensional. To this end, a light field contains not only colour information for each scene point but as well a directional information. It allows in post-processing, for example, refocusing, view change, depth-based colour grading, alpha-mating, and even relighting, assuming diffuse surfaces.

This thesis addresses the problems of an HDR content capture and its calibration, which allows to obtain the actual luminance values of a scene, thus making a step towards a more realistic scene representation and recreation.

Moreover, we address a challenging problem of editing angularly varying effects in a captured content. Hence, we make a step towards not only being able to capture a real scene but also modify it according, for example, to a certain artistic intention. Both, an HDR content and a light-field content, provide a large capacity for rendering diverse perceptual and artistic effects.

In the following Sections 7.1.1–7.1.3 we give a summary of each of core chapters of this thesis and possible future work within each of individual research topics.

7.1.1 HDR Video Capturing

A huge mobile phone market means that an enormous number of images and videos originate from mobile devices. Our research, presented in Chapter 4, is oriented toward improving their visual quality by capturing higher dynamic range than such cameras are natively capable off, while minimizing motion artefacts. To this end, a novel real-time metering approach is combined with off-line HDR video reconstruction. Our hypothesis, backed by a psychophysical experiment (Section 4.1), is that failing to capture the full dynamic range of a scene is a lesser evil than introducing motion artefacts that are seen as ghosting. The proposed metering algorithm is therefore designed to seamlessly adapt to the scene that is being captured: if sufficient motion warrants it, moving objects are captured with high fidelity; otherwise the scene's dynamic range is captured with high fidelity. This allows the algorithm to work within the potential hardware constraints imposed by mobile devices (sensor dynamic range, frame-rate and computing constraints). The resulting stream of exposures is then combined into a full HDR sequence in an off-line process that is well matched to our metering algorithm.

Analysis of our results shows that our capture strategy yields HDR video that is visually superior to the current state-of-the-art for all sequences that we have available (Section 4.5). While we achieve noticeable improvement in dynamic range without introducing disturbing artefacts, our solution inherited a drawback of all previous work on multi-exposure video reconstruction. Namely, the reconstruction phase in multi-exposure techniques has to be pushed to real-time performance in order to gain popularity. This goal is facilitated by recent progress in motion estimation and correspondence estimation algorithms, as well as texture synthesis.

7.1.2 Calibration

In Chapter 5 we presented an automatic algorithm to estimate absolute pixel values from single HDR images, using only the EXIF header information as well as GPS data. Note that, SLR cameras nowadays start to have built-in GPS systems and all recent mobile phones provide this tool. Conversely, the development of HDR technologies on mobile phones is a recent trend that makes our work increasingly relevant.

For images containing the full sky dome, as would be the case with light probes used for image-based lighting, our approach enables the estimation of absolute data without requiring any additional information.

Our method is targeted at outdoor scenes with HDR illumination, which are analysed to recover camera parameters as well as the turbidity of the atmosphere. Using a sky luminance model, the relative zenith luminance is estimated and subsequently correlated with the absolute zenith luminance calculated from GPS data. We compared several zenith luminance models and developed a trigonometric interpolation between two of them to obtain higher accuracy over a wider range of sun elevation angles. Additionally, we performed a comparison of the well-known Perez sky model with its recent extension [Hosek and Wilkie 2012] in the context of our work. The algorithm was validated using an existing HDR image dataset as well as our own calibrated and annotated HDR image collection.

In Section 5.4.1 we demonstrated that for scenes with high dynamic range simple EXIF-based approaches can be inaccurate. Limited to the images containing sky regions, our algorithm provides the solution which does not require any measurement equipment and can be used for certain existing imaginary. Although, sky appearance is a highly complicated phenomenon and cluttered skies will

be the difficult case for our algorithm, our algorithm is sufficiently robust as tested on the images with different sky types.

Absolute values are required when one is interested in rendering a certain perception effect. By its nature, HDR imaging can capture luminance range covering the scotopic, mesopic and photopic vision. This corresponds to different perception of colours. For example, in dim conditions cone cells do not function as well as rod cells. Thus, scotopic vision happens completely through rod cells, which are most sensitive to wavelengths of light on the electromagnetic spectrum of 498nm, which is the blue-green colours. This leads to a loss of colour vision. An example usage of our approach in appearance reproduction is given in the Figure 7.1. Further, with the advent of



a. Unscaled input | b. Our scaling | c. Ground truth scaling | d. CIE ΔE94 (a vs. c) | e. CIE ΔE94 (b vs. c)

Figure 7.1: An unscaled HDR image and scaled with our estimate was processed using an appearance model [Reinhard et al. 2012]. The CIE ΔE94 differences shows good correspondence between our estimate and the ground-truth.

scene-referred HDR displays, we think that there will be an increasing demand for images that are not only calibrated (i.e. linearized) but also absolute. Our algorithm helps to create such data, and can therefore be seen as a tool that turns cameras into high resolution light meters.

7.1.3 Gloss Editing

A new technique from Chapter 6 allows to modify glossy appearance, and in particular apparent roughness, of materials in light fields. It is based on the observation that plausible material edits do not require a complete scene reconstruction which is usually expensive and prone to artefacts. Instead, it is sufficient to consistently propagate existing specular information in the spatial and angular domains. For this purpose, we have developed a novel filtering approach that is applied to the specular component of the light field.

Our method does not require explicit information about geometry, material, or illumination: During our filtering step, we exploit the fact that such information is encoded in the light field implicitly. As a preprocessing step, we compute depth information and a specular-diffuse separation that will be used in the filtering step. With respect to the recent work by Boyadzhiev et al. [2015], our technique provides different edits. While they modify high level properties, such as oiliness, glow or blemishes of the skin, we focus on the editing of glossy appearance, and roughness in particular. Furthermore, our filtering allows to propagate edits to views of the light field which had no gloss before, while, for instance, the approach by Boyadzhiev et al. is designed for 2D images, and cannot support such operations in a trivial manner.

Our technique has been evaluated on several scenes. It as was compared to ground-truth renderings, by optimizing filtering parameters to obtain results matching the ground-truth results.

Currently, our algorithm is only capable of increasing surface roughness. The opposite operation is an interesting avenue for future work.

To further improve an accuracy of the method one might need to account for local curvature

and normalize the kernel accordingly. The idea behind is that the image of a surface with high curvature contains a much larger number of samples than a surface with low curvature. This might lead to uneven energy distribution. Extending further the normalisation strategy one can extend our approach for unstructured input, such as dense video stream. In the future, perceptual experiments could be used to further evaluate our results and investigate if the technique can be further simplified.

7.2 Future Work

Capturing a high-dynamic-range light field is an interesting and challenging problem as already was discussed in Section 2.7. Specifically, when a plenoptic camera is used for light-field capture, multiple views are stored on a single sensor. This means that the sensor size already has to be large enough to provide sufficient pixel resolution for each of individual views. As discussed in Section 2.3, that imposes the limitation on the captured dynamic range, which requires larger pixel sizes and consequently even a larger sensor size. Next, when a multiple camera setup is used, in order to keep the cost of the setup reasonable, each of individual cameras has to be affordable.

One feasible solution for capturing an HDR light field was suggested by Schedl et al. [2014]. This approach is limited to static scenes and does not universally solve the problem of motion blur in the frames with longer exposure times. There is a huge potential in borrowing the ideas from the single-view-HDR-capturing techniques and extending them to light fields. For example, one can approach the problem of reconstruction an HDR light field from the input similar to the one used in [Schedl et al. 2014], in an absolutely different way, by considering EPIs. The EPIs of such a light field will have an interlaced structure. Hence, it is natural to borrow ideas from deinterlacing techniques for multi-exposure sensors [Heide et al. 2014, Cho et al. 2014] or decoding techniques [Serrano et al. 2016] to perform a reconstruction of an HDR light field from the interlaced EPIs.

Another interesting outcome of intersecting light-field and HDR techniques can be in considering the problem of correspondence estimation in an LDR sequence of differently exposed frames as a problem of depth/point cloud estimation [Klose et al. 2015, Wolff et al. 2016, Yücer et al. 2016]. This, for example, potentially can improve temporal coherence in reconstructed overexposed regions.

Finally, HDR light fields facilitate work with non-Lambertian objects. The rich information about specular regions encoded in an HDR light field (e.g., Figure 2.8) can be used for depth estimation from the specular shapes in EPIs, for specular separation or BRDF estimation. For example, recent work on specular separation in light fields [Sulc et al. 2016] can benefit from the HDR-light-field input.

In conclusion, the work in this thesis makes important steps towards being able to capture and recreate a scene, as well as edit the captured content in a realistic way.

Bibliography (Own Work)

Gryaditskaya, Y., Pouli, T., Reinhard, E., and Seidel, H.-P. 2014. Sky based light metering for high dynamic range images. *Computer Graphics Forum*, 33(7), 61–69.

Gryaditskaya, Y., Pouli, T., Reinhard, E., Myszkowski, K., and Seidel, H.-P. 2015. Motion aware exposure bracketing for HDR video. *Computer Graphics Forum*, 34(4), 119–130.

Gryaditskaya, Y., Masia, B., Didyk, P., Myszkowski, K., and Seidel, H.-P. 2016. Gloss editing in light fields. In *VMV'16: International Workshop on Vision, Modeling and Visualization*.

Pouli, F. T., Reinhard, E., and Gryaditskaya, Y. January 19 2016. Method for generating an hdr image of a scene based on a tradeoff between brightness distribution and motion. US Patent App. 15/001,121.

Bibliography

2006. Photography - digital still cameras - determination of exposure index, iso speed ratings, standard output sensitivity, and recommended exposure index.

A1EX. July 2013. Dynamic range improvement for some canon dSLRs by alternating ISO during sensor readout. Technical report, A1EX.

Ackermann, J. and Goesele, M. 2013. How bright is the moon? Recovering and using absolute luminance values from internet images. In *Proc. CCIW*, pages 90–104, Berlin, Heidelberg. Springer-Verlag.

Adelson, E. H. and Bergen, J. R. 1991. The plenoptic function and the elements of early vision. In *Proc. Computational Models of Visual Processing*, pages 3–20. MIT Press.

Aggarwal, M. and Ahuja, N. 2004. Split aperture imaging for high dynamic range. *INT J COMPUT VISION*, 58(1), 7–17.

Aguerrebere, C., Almansa, A., Gousseau, Y., Delon, J., and Musé, P. 2014. Single shot high dynamic range imaging using piecewise linear estimators. In *Proc. IEEE ICCP*, pages 1–10.

An, X. and Pellacini, F. 2008. AppProp: all-pairs appearance-space edit propagation. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 27, 40:1–40:9.

Artusi, A., Banterle, F., and Chetverikov, D. 2011. A survey of specularity removal methods. *Comp. Graph. Forum*, 30(8), 2208–2230.

Barakat, N., Hone, A. N., and Darcie, T. E. 2008. Minimal-bracketing sets for high-dynamic-range image capture. *IEEE Trans. Image Process.*, 17(10), 1864–1875.

Beck, J. and Prazdny, S. 1981. Highlights and the perception of glossiness. *Perception & Psychophysics*, 30(4), 407–410.

Ben-Artzi, A., Overbeck, R., and Ramamoorthi, R. 2006. Real-time BRDF editing in complex lighting. *ACM Trans. Graph.*, 25(3), 945–954.

Berger, K., Lipski, C., Linz, C., Sellent, A., and Magnor, M. 2010. A ghosting artifact detector for interpolated image quality assessment. In *Proc. IEEE ISCE*, pages 1–6.

Blake, A. and Bülthoff, H. 1990. Does the brain know the physics of specular reflection? *Nature*, 343(6254), 165–168.

Bolles, R. C., Baker, H. H., and Marimont, D. H. 1987. Epipolar-plane image analysis: An approach to determining structure from motion. *INT J COMPUT VISION*, 1(1), 7–55.

Bouzaraa, F., Halfaoui, I., and Urfalioglu, O. September 2016. Ghost-free dual-exposure hdr for dynamic scenes. In *Proc. IEEE ICIP*, pages 1739–1743.

Boyadzhiev, I., Bala, K., Paris, S., and Adelson, E. 2015. Band-sifting decomposition for image-based material editing. *ACM Trans. Graph.*, 34(5), 163.

Brainard, D. H. 1997. The psychophysics toolbox. *Spatial vision*, 10, 433–436.

Brajovic, V. and Kanade, T. 1996. A sorting image sensor: An example of massively parallel intensity-to-time processing for low-latency computational sensors. In *Proc. IEEE International Conference on Robotics and Automation*, volume 2, pages 1638–1643.

Brajovic, V. M., Miyagawa, R., and Kanade, T. 1998. Temporal photoreception for adaptive dynamic range image sensing and encoding. *Neural Networks*, 11(7), 1149–1158.

Chadwick, A. and Kentridge, R. 2015. The perception of gloss: A review. *Vision Research*, 109, Part B, 221 – 235. Perception of Material Properties (Part I).

Chen, B., Ofek, E., Shum, H.-Y., and Levoy, M. 2005. Interactive deformation of light fields. In *Proc. I3D*, pages 139–146.

Chen, C., Lin, H., Yu, Z., Kang, S. B., and Yu, J. 2014. Light field stereo matching using bilateral statistics of surface cameras. In *Proc. IEEE CVPR*, pages 1518–1525.

Cho, H., Kim, S. J., and Lee, S. 2014. Single-shot high dynamic range imaging using coded electronic shutter. *Comp. Graph. Forum*, 33(7), 329–338.

Chupeau, B. 1993. Multiscale motion estimation. In *Proc. International Workshop on HDTV*, pages 233–240.

CIE. 2003. *Spatial distribution of daylight — CIE standard general sky*. CIE Standard S 011/E:2003CIE Central Bureau, Vienna.

Colbert, M. and Pattanaik, S. 2006. BRDF-shop: Creating physically correct bidirectional reflectance distribution functions. *IEEE Comput. Graph. Appl.*, pages 30–36.

Criminisi, A., Kang, S. B., Swaminathan, R., Szeliski, R., and Anandan, P. 2005. Extracting layers and analyzing their specular properties using epipolar-plane-image analysis. *Comput. Vis. Image. Und.*, 97(1), 51–85.

Daly, S., Xu, N., Crenshaw, J., and Zunjarrao, V. J. 2015. A psychophysical study exploring judder using fundamental signals and complex imagery. *SMPTE Mot. Imag. J*, 124(7), 62–70.

Debevec, P. 2002. Image-based lighting. *IEEE COMPUT GRAPH*, 22(2), 26–34.

Debevec, P. 2008. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proc. ACM SIGGRAPH 2008 classes*, page 32. ACM.

Debevec, P. E. and Malik, J. 1997. Recovering high dynamic range radiance maps from photographs. In *Proc. ACM SIGGRAPH*, pages 369–378.

Doerschner, K., Fleming, R. W., Yilmaz, O., Schrater, P. R., Hartung, B., and Kersten, D. 2011. Visual motion and the perception of surface material. *Current Biology*, 21(23), 2010–2016.

Dror, R. O., Adelson, E. H., and Willsky, A. S. 2001. Recognition of surface reflectance properties from a single image under unknown real-world illumination. In *Proc. IEEE Workshop on Identifying Objects Across Variations in Lighting*, pages 1–8.

Efrat, N., Didyk, P., Foshey, M., Matusik, W., and Levin, A. 2016. Cinema 3D: Large scale automultiscopic display. *ACM Trans. Graph.*, 35(4), 59:1–59:12.

Eisemann, M., Frahm, J.-M., Remion, Y., and Ismael, M. *Reconstruction of Dense Correspondences*, chapter 8, pages 113–133. CRC Press, May 2015. ISBN 9781482243819.

Fairchild, M. D. 2007. The HDR photographic survey. In *Proc. Color and Imaging Conference*, volume 2007, pages 233–238. Society for Imaging Science and Technology.

Fairchild, M. D. 2013. *Color appearance models*. John Wiley & Sons.

Felzenszwalb, P. F. and Huttenlocher, D. P. 2004. Efficient graph-based image segmentation. *INT J COMPUT VISION*, 59(2), 167–181.

Ferwerda, J. A. 2001. Elements of early vision for computer graphics. *IEEE Comput. Graph. Appl.*, 21(5), 22–33.

Fleming, R. 2012. Human perception: Visual heuristics in the perception of glossiness. *Current Biology*, 22(20), R865–R866.

Fleming, R., Torralba, A., and Adelson, E. 2004. Specular reflections and the perception of shape. *J. Vis.*, 4(9), 798–820.

Fleming, R. W., Dror, R. O., and Adelson, E. H. 2003. Real-world illumination and the perception of surface reflectance properties. *J. Vis.*, 3, 347–368.

Fliegel, K. and Havlin, J. 2009. Imaging photometer with a non-professional digital camera. In *Proc. SPIE*, volume 7443.

Froehlich, J., Grandinetti, S., Eberhardt, B., Walter, S., Schilling, A., and Brendel, H. 2014. Creating cinematic wide gamut HDR-video for the evaluation of tone mapping operators and HDR-displays. In *Proc. SPIE*, volume 9023.

Furukawa, Y., Hernández, C., et al. 2015. *Multi-view stereo: A tutorial*. Now Publishers Incorporated.

Gallo, O., Tico, M., Manduchi, R., Gelfand, N., and Pulli, K. 2012. Metering for exposure stacks. *Comp. Graph. Forum*, 31(2), 479–488.

Gershun, A. 1936. "The light field", Moscow, 1936. *Translated by P. Moon and G. Timoshenko in Journal of Mathematics and Physics, 1939*, 18(1), 51–151.

Ginosar, R. and Gnusin, A. 1997. A wide dynamic range CMOS image sensor. In *Proc. IEEE Workshop on CCD and Advanced Image Sensors*, volume 32.

Goldluecke, B., Klehm, O., Wanner, S., and Eisemann, E. *Plenoptic Cameras*, pages 65–77. CRC Press, 2015.

Goldluecke, B. and Wanner, S. 2013. The variational structure of disparity and regularization of 4D light fields. In *Proc. IEEE CVPR*, pages 1003–1010.

Gortler, S. J., Grzeszczuk, R., Szeliski, R., and Cohen, M. F. 1996. The lumigraph. In *Proc. ACM SIGGRAPH*, pages 43–54.

Granados, M., Ajdin, B., Wand, M., Theobalt, C., Seidel, H.-P., and Lensch, H. P. A. 2010. Optimal HDR reconstruction with linear digital cameras. In *Proc. IEEE CVPR*, pages 215–222.

Grossberg, M. D. and Nayar, S. K. 2003. High dynamic range from multiple images: Which exposures to combine? In *Proc. ICCV Workshop on CPMCV*.

Gryaditskaya, Y., Pouli, T., Reinhard, E., and Hans-Peter, S. 2014. Calibrated HDR dataset. https://people.mpi-inf.mpg.de/~jgryadit/Papers/LightMetering/index.html#pg14.

Gu, J., Hitomi, Y., Mitsunaga, T., and Nayar, S. 2010. Coded rolling shutter photography: Flexible space-time sampling. In *Proc. IEEE ICCP*, pages 1–8.

Gupta, M., Iso, D., and Nayar, S. K. December 2013. Fibonacci exposure bracketing for high dynamic range imaging. In *Proc. IEEE ICCV*.

Guthier, B., Kopf, S., and Effelsberg, W. 2013. Algorithms for a real-time HDR video system. *Pattern. Recogn. Lett.*, 34(1), 25–33.

Gutierrez, D., Seron, F. J., Lopez-Moreno, J., Sanchez, M. P., Fandos, J., and Reinhard, E. 2008. Depicting procedural caustics in single images. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, 27 (5), 120:1–9.

Haber, T., Fuchs, C., Bekaert, P., Seidel, H.-P., Goesele, M., and Lensch, H. 2009. Relighting objects from image collections. In *Proc. CVPR*, pages 627–634. IEEE.

Hajisharif, S., Kronander, J., and Unger, J. 2015. Adaptive dualISO HDR reconstruction. *EURASIP Journal on Image and Video Processing*, 2015(1), 1.

Hasinoff, S. W., Durand, F., and Freeman, W. T. 2010. Noise-optimal capture of high dynamic range photography. In *Proc. IEEE CVPR*, pages 553–560.

He, K., Sun, J., and Tang, X. 2010. Guided image filtering. In *Proc. ECCV*, pages 1–14, Berlin, Heidelberg. Springer-Verlag.

Heber, S. and Pock, T. 2014. Shape from light field meets robust pca. In *Proc. ECCV*, pages 751–767. Springer.

Heber, S. and Pock, T. 2016. Convolutional networks for shape from light field. In *Proc. IEEE CVPR*, pages 3746–3754.

Heber, S., Ranftl, R., and Pock, T. 2013. Variational shape from light field. In *Proc. International Workshop EMMCVPR*, pages 66–79, New York, NY, USA. Springer-Verlag New York, Inc.

Heide, F., Steinberger, M., Tsai, Y.-T., Rouf, M., Pajak, D., Reddy, D., Gallo, O., Liu, J., Heidrich, W., Egiazarian, K., Kautz, J., and Pulli, K. 2014. FlexISP: A flexible camera image processing framework. *ACM Trans. Graph.*, 33(6).

Hirakawa, K. 2010. Optimal exposure control for high dynamic range imaging. In *Proc. IEEE ICIP*, pages 3137–3140.

Hirakawa, K. and Simon, P. M. 2011. Single-shot high dynamic range imaging with conventional camera hardware. In *Proc. IEEE ICCV*, pages 1339–1346. IEEE.

Hirsch, M., Wetzstein, G., and Raskar, R. 2014. A compressive light field projection system. *ACM Trans. Graph.*, 33(4), 58.

Hoefflinger, B. 2007. *High-dynamic-range (HDR) vision*. Springer.

Hoiem, D., Efros, A. A., and Hebert, M. 2007. Recovering surface layout from an image. *INT J COMPUT VISION*, 75(1), 151–172.

Hosek, L. and Wilkie, A. July 2012. An analytic model for full spectral sky-dome radiance. *ACM Trans. Graph.*, 31(4), 95:1–95:9.

Hu, J., Gallo, O., Pulli, K., and Sun, X. 2013. HDR deghosting: How to deal with saturation? In *Proc. IEEE CVPR*, pages 1163–1170.

Ilstrup, D. and Manduchi, R. 2010. One-shot optimal exposure control. In *Proc. ECCV*, pages 200–213.

Ineichen, P. and Molineaux, B. 1994. Sky luminance data validation: Comparison of 7 models with 4 data banks. *Solar Energy*, 52(4), 337–346.

ITU-R. 2012. Recommendation ITU-R BT.500-13: Methodology for the subjective assessment of the quality of television pictures. Technical report, ITU-R.

Jarabo, A., Masia, B., and Gutierrez, D. 2011. Efficient propagation of light field edits. In *Proc. SIACG*, pages 75–80.

Jarabo, A., Masia, B., Bousseau, A., Pellacini, F., and Gutierrez, D. 2014. How do people edit light fields? *ACM Trans. Graph. (Proc. SIGGRAPH)*, 33(4).

Kalantari, N. K., Shechtman, E., Barnes, C., Darabi, S., Goldman, D. B., and Sen, P. 2013. Patch-based high dynamic range video. *ACM Trans. Graph.*, 32(6), 202:1–202:8.

Kamal, M. H., Favaro, P., and Vandergheynst, P. 2015. A convex solution to disparity estimation from light fields via the primal-dual method. In *Proc. International Workshop EMMCVPR*, pages 350–363. Springer.

Kang, S. B., Uyttendaele, M., Winder, S., and Szeliski, R. 2003. High dynamic range video. *ACM Trans. Graph.*, pages 319–325.

Karayel, M., Navvab, M., Neeman, E., and Selkowitz, S. E. 1984. Zenith luminance and sky luminance distributions for daylighting calculations. *Energy And Buildings*, 6(3), 283–291.

Karsch, K., Hedau, V., Forsyth, D., and Hoiem, D. 2011. Rendering synthetic objects into legacy photographs. *ACM Trans. Graph.*, 30(6), 157:1–157:12.

Kawakami, R., Zhao, H., Tan, R. T., and Ikeuchi, K. 2013. Camera spectral sensitivity and white balance estimation from sky images. *INT J COMPUT VISION*, 105(3), 187–204.

Keinert, J., Wetzel, M., Schöberl, M., Schäfer, P., Zilly, F., Bätz, M., Fößel, S., and Kaup, A. 2014. Cost-effective multi-camera array for high quality video with very high dynamic range. In *Proc. IS&T/SPIE Electronic Imaging*, pages 90230Y–90230Y. International Society for Optics and Photonics.

Khan, E. A., Reinhard, E., Fleming, R. W., and Bülthoff, H. H. 2006. Image-based material editing. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 25(3), 654–663.

Kim, C., Zimmer, H., Pritch, Y., Sorkine-Hornung, A., and Gross, M. H. 2013. Scene reconstruction from high spatio-angular resolution light fields. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 32(4), 73–1.

Kittler, R., Perez, R., and Darula, S. 1998. *A set of standard skies characterizing daylight conditions for computer and energy conscious design*. US SK 92 052 Final Report, CA SAS Bratislava.

Klose, F., Wang, O., Bazin, J.-C., Magnor, M., and Sorkine-Hornung, A. 2015. Sampling based scene-space video processing. *ACM Trans. Graph.*, 34(4), 67.

Kronander, J., Gustavson, S., Bonnet, G., and Unger, J. 2013. Unified HDR reconstruction from raw CFA data. In *Proc. IEEE ICCP*, pages 1–9.

Lalonde, J.-F., G., N. S., and Efros, A. A. 2010. What do the sun and the sky tell us about the camera? *INT J COMPUT VISION*, 88(1), 24–51.

Lawrence, J., Ben-Artzi, A., DeCoro, C., Matusik, W., Pfister, H., Ramamoorthi, R., and Rusinkiewicz, S. 2006. Inverse shade trees for non-parametric material representation and editing. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 25(3), 735–745.

Levin, A., Weiss, Y., Durand, F., and Freeman, W. T. 2011. Efficient marginal likelihood optimization in blind deconvolution. In *Proc. IEEE CVPR*, pages 2657–2664. IEEE.

Levoy, M. and Hanrahan, P. 1996. Light field rendering. In *Proc. ACM SIGGRAPH*, pages 31–42.

Li, C. and Zhang, X. 2015. High dynamic range and all-focus image from light field. In *Proc. IEEE CIS and IEEE RAM*, pages 7–12. IEEE.

Lin, H., Chen, C., Bing Kang, S., and Yu, J. 2015. Depth recovery from light field using focal stack symmetry. In *Proc. IEEE ICCV*, pages 3451–3459.

Lin, H.-Y. and Chang, W.-Z. 2009. High dynamic range imaging for stereoscopic scene representation. In *Proc. ICIP*, pages 4305–4308.

Lin, S., Li, Y., Kang, S. B., Tong, X., and Shum, H.-Y. 2002. Diffuse-specular separation and depth recovery from image sequences. In *Proc. ECCV*, pages 210–224. Springer.

Lopez-Moreno, J., Sundstedt, V., Sangorrin, F., and Gutierrez, D. 2010. Measuring the perception of light inconsistencies. In *Proc. APGV*. ACM Press.

Madsen, C. B. and Nielsen, M. 2008. Towards probe-less augmented reality - a position paper. In *Proc. VISIGRAPP*, pages 255–261.

Manakov, A., Restrepo, J. F., Klehm, O., Hegedüs, R., Eisemann, E., Seidel, H.-P., and Ihrke, I. 2013. A reconfigurable camera add-on for high dynamic range, multi-spectral, polarization, and light-field imaging. *ACM Trans. Graph.*, 32(4), 47:1–47:14.

Mann, S. and Picard, R. 1995. On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proc. IS&T*, pages 442–448.

Mantiuk, R., Daly, S., and Kerofsky, L. 2008. Display adaptive tone mapping. *ACM Trans. Graph.*, 27(3), 68.

Mantiuk, R. K., Myszkowski, K., and Seidel, H.-P. 2015. *High dynamic range imaging*. Wiley Online Library.

Marlow, P. J., Kim, J., and Anderson, B. L. 2012. The perception and misperception of specular surface reflectance. *Current Biology*, 22(20), 1909–1913.

Masia, B., Agustin, S., Fleming, R. W., Sorkine, O., and Gutierrez, D. December 2009. Evaluation of reverse tone mapping through varying exposure conditions. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, 28(5), 160:1–160:8. ISSN 0730-0301.

McCartney, E. J. 1976. *Optics of the Atmosphere: Scattering by molecules and particles*. John Wiley & Sons, New York.

Mertens, T., Kautz, J., and Reeth, F. V. 2007. Exposure fusion. *Comp. Graph. Forum (Proc. Pacific Graphics)*, pages 382–390.

Mihara, H., Funatomi, T., Tanaka, K., Kubo, H., Nagahara, H., and Mukaigawa, Y. 2016. 4D light field segmentation with spatial and angular consistencies. In *Proc. IEEE CVPR*.

Mitsunaga, T. and Nayar, S. K. 1999. Radiometric self calibration. In *Proc. IEEE CVPR*, pages 374–380.

Moroney, N., Fairchild, M. D., Hunt, R. W., Li, C., Luo, M. R., and Newman, T. 2002. The CIECAM02 color appearance model. In *Proc. Color and Imaging Conference*, volume 2002, pages 23–27. Society for Imaging Science and Technology.

Narasimha, R. and Batur, A. U. June 2012. Method and apparatus for producing sharp frames with less blur. US Patent 8,194,141.

Nayar, S. K., Branzoi, V., and Boult, T. E. June 2004. Programmable imaging using a digital micromirror array. In *Proc. IEEE CVPR*, volume 1, pages I–436–I–443 Vol.1. doi: 10.1109/CVPR.2004.1315065.

Nayar, S. K. and Branzoi, V. 2003. Adaptive dynamic range imaging: Optical control of pixel exposures over space and time. In *Proc. IEEE ICCV*, pages 1168–1175.

Nayar, S. K. and Mitsunaga, T. 2000. High dynamic range imaging: Spatially varying pixel exposures. In *Proc. IEEE CVPR*, volume 1, pages 472–479. IEEE.

Nayar, S. K., Branzoi, V., and Boult, T. E. 2006. Programmable imaging: Towards a flexible camera. *INT J COMPUT VISION*, 70(1), 7–22.

Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M., and Hanrahan, P. 2005. Light field photography with a hand-held plenoptic camera. *Comp. Science TR*, 2(11), 1–11.

Ngan, A., Durand, F., and Matusik, W. 2006. Image-driven navigation of analytical BRDF models. In *Proc. EGSR*.

Nguyen, C. H., Scherzer, D., Ritschel, T., and Seidel, H.-P. 2013. Material editing in complex scenes by surface light field manipulation and reflectance optimization. *Comp. Graph. Forum (Proc. EUROGRAPHICS)*, 32(2).

Oh, B. M., Chen, M., Dorsey, J., and Durand, F. 2001. Image-based modeling and photo editing. In *Proc. ACM SIGGRAPH*, pages 433–442.

Ostrovsky, Y., Cavanagh, P., and Sinha, P. 2005. Perceiving illumination inconsistencies in scenes. *Perception*, 34(11).

Pellacini, F. and Lawrence, J. 2007. Appwand: Editing measured materials using appearance-driven optimization. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 26(3).

Perez, R., Seals, R., and Michalsky, J. 1993. All-weather model for sky luminance distribution-preliminary configuration and validation. *Solar Energy*, 50(3), 235–245.

Poynton, C. A. 1996. *A technical introduction to digital video*. John Wiley & Sons, Inc.

Preetham, A. J., Shirley, P., and Smits, B. 1999. A practical analytic model for daylight. In *Proc. ACM SIGGRAPH*, pages 91–100. ACM Press/Addison-Wesley Publishing Co.

Reda, I. and Andreas, A. 2004. Solar position algorithm for solar radiation applications. *Solar Energy*, 76(5), 577–589.

Reinhard, E., Heidrich, W., Debevec, P., Pattanaik, S., Ward, G., and Myszkowski, K. 2010. *High Dynamic Range Imaging: Acquisition, Display and Image-Based Lighting*. Morgan Kaufmann, San Francisco, 2nd edition.

Reinhard, E., Pouli, T., Kunkel, T., Long, B., Ballestad, A., and Damberg, G. 2012. Calibrated image appearance reproduction. *ACM Trans. Graph.*, 31(6), 201.

Reinhard, E. and Devlin, K. 2005. Dynamic range reduction inspired by photoreceptor physiology. *IEEE Trans. Vis. Comput. Graphics*, 11(1), 13–24.

Ritschel, T., Okabe, M., Thormählen, T., and Seidel, H.-P. 2009. Interactive reflection editing. *ACM Trans. Graph.*, 28(5).

Robertson, M. A., Borman, S., and Stevenson, R. L. 2003. Estimation-theoretic approach to dynamic range enhancement using multiple exposures. *J Electronic Imaging*, 12(2), 219–228.

Schedl, D. C., Birklbauer, C., and Bimber, O. 2014. Coded exposure HDR light-field video recording. *Comp. Graph. Forum*, 33(2), 33–42.

Schmidt, S. and Li, Z.-N. 2011. High dynamic range stereo video using sift and simultaneous multi-exposure. In *Proc. IEEE ICSIPA*, pages 82–87. IEEE.

Schoberl, M., Belz, A., Seiler, J., Foessel, S., and Kaup, A. 2012. High dynamic range video by spatially non-regular optical filtering. In *Proc. IEEE ICIP*.

Seger, U., Apel, U., and Höfflinger, B. 1999. HDRC-imagers for natural visual perception. *Handbook of Computer Vision and Application*, 1, 223–235.

Seitz, S. M. and Kutulakos, K. N. 2002. Plenoptic image editing. *INT J COMPUT VISION*, 48(2), 115–129.

Sen, P., Kalantari, N. K., Yaesoubi, M., Darabi, S., Goldman, D. B., and Shechtman, E. 2012. Robust patch-based HDR reconstruction of dynamic scenes. *ACM Trans. Graph.*, 31(6), 203:1–203:11.

Serrano, A., Heide, F., Gutierrez, D., Wetzstein, G., and Masia, B. 2016. Convolutional sparse coding for high dynamic range imaging. *Comp. Graph. Forum*, 35(2), 153–163.

Sheng, H., Zhang, S., Zhu, G., and Xiong, Z. Sept 2015. Guided integral filter for light field stereo matching. In *Proc. IEEE ICIP*, pages 852–856.

Simakov, D., Caspi, Y., Shechtman, E., and Irani, M. 2008. Summarizing visual data using bidirectional similarity. In *Proc. IEEE CVPR*.

Soler, A. and Gopinathan, K. K. 2000. A study of zenith luminance on Madrid cloudless skies. *Solar Energy*, 69(5), 403–411.

Soler, A. and Gopinathan, K. K. 2001. Analysis of zenith luminance data for all sky conditions. *Renewable Energy*, 24(2), 185–196.

Soler, A., Gopinathan, K. K., and Clarosa, S. 2001. A study on zenith luminance on Madrid overcast skies. *Renewable Energy*, 23(1), 49–55.

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

Stengel, M., Bauszat, P., Eisemann, M., Eisemann, E., and Magnor, M. 2015. Temporal video filtering and exposure control for perceptual motion blur. *IEEE Trans. Vis. Comput. Graphics*, 21 (3), 1–1.

Stich, T., Tevs, A., and Magnor, M. A. 2006. Global depth from epipolar volumes–a general framework for reconstructing non-lambertian surfaces. In *Proc. 3DPVT*, pages 916–923. Citeseer.

Stumpfel, J., Tchou, C., Jones, A., Hawkins, T., Wenger, A., and Debevec, P. 2004. Direct HDR capture of the sun and sky. In *Proc. Afrigraph*, pages 145–149. ACM.

Sulc, A., Alperovich, A., Marniok, N., and Goldluecke, B. 2016. Reflection separation in light fields based on sparse coding and specular flow. In *Proc. VMV*.

Talton, J. O., Gibson, D., Yang, L., Hanrahan, P., and Koltun, V. 2009. Exploratory modeling with collaborative design spaces. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, 28(5), 167:1–10.

Tao, L., Yuan, L., and Sun, J. 2009. Skyfinder: attribute-based sky image search. *ACM Trans. Graph.*, 28(3), 68.

Tao, M., Su, J.-C., Wang, T.-C., Malik, J., and Ramamoorthi, R. 2015a. Depth estimation and specular removal for glossy surfaces using point and line consistency with light-field cameras. *IEEE TPAMI*, (1).

Tao, M. W., Srinivasan, P. P., Malik, J., Rusinkiewicz, S., and Ramamoorthi, R. 2015b. Depth from shading, defocus, and correspondence using light-field angular coherence. In *Proc. CVPR*, pages 1940–1948.

Tao, M. W., Su, J.-C., Wang, T.-C., Malik, J., and Ramamoorthi, R. 2016. Depth estimation and specular removal for glossy surfaces using point and line consistency with light-field cameras. *IEEE TPAMI*, 38(6), 1155–1169.

Tocci, M. D., Kiser, C., Tocci, N., and Sen, P. 2011. A versatile HDR video production system. *ACM Trans. Graph.*, 30(4), 41:1–41:10.

Tourapis, A. M. 2002. Enhanced predictive zonal search for single and multiple frame motion estimation. In *Proc. SPIE*, volume 4671, pages 1069–1079.

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

Urban, F., Nezan, J.-F., and Raulet, M. 2009. HDS, a real-time multi-DSP motion estimator for MPEG-4 H.264 AVC high definition video encoding. *Journal of Real-Time Image Processing*, 4 (1), 23–31.

Vangorp, P., Laurijssen, J., and Dutré, P. 2007. The influence of shape on the perception of material reflectance. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 26(3).

Vangorp, P., Chaurasia, G., Laffont, P.-Y., Fleming, R., and Drettakis, G. 2011. Perception of visual artifacts in image-based rendering of façades. *Comp. Graph. Forum*, 30(4), 1241–1250.

Venkataraman, K., Lelescu, D., Duparré, J., McMahon, A., Molina, G., Chatterjee, P., Mullis, R., and Nayar, S. 2013. PiCam: An ultra-thin high performance monolithic camera array. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, 32(6), 166:1–166:13.

Vergne, R., Barla, P., Fleming, R. W., and Granier, X. 2012. Surface flows for image-based shading design. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 31(4), 94:1–94:9.

Vetro, A., Tourapis, A., Muller, K., and Chen, T. 2011. 3D-TV content storage and transmission. *IEEE Trans. on Broadcasting*, 57(2), 384–394.

Wan, G., Li, X., Agranov, G., Levoy, M., and Horowitz, M. 2012. CMOS image sensors with multi-bucket pixels for computational photography. *IEEE J. Solid-State Circuits*, 47(4), 1031–1042.

Wang, L., Lin, S., Lee, S., Guo, B., and Shum, H.-Y. 2005. Light field morphing using 2D features. *IEEE TVCG*, 11(1), 25–34. ISSN 1077-2626.

Wang, T.-C., Efros, A. A., and Ramamoorthi, R. 2015. Occlusion-aware depth estimation using light-field cameras. In *Proc. IEEE ICCV*, pages 3487–3495.

Wang, T.-C., Srikanth, M., and Ramamoorthi, R. 2016. Depth from semi-calibrated stereo and defocus. In *Proc. IEEE CVPR*, pages 3717–3726.

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

Wanner, S. and Goldluecke, B. 2012. Globally consistent depth labeling of 4D light fields. In *Proc. IEEE CVPR*, pages 41–48. IEEE.

Wanner, S. and Goldluecke, B. 2014. Variational light field analysis for disparity estimation and super-resolution. *IEEE TPAMI*, 36(3), 606–619.

Wanner, S., Straehle, C., and Goldluecke, B. 2013. Globally consistent multi-label assignment on the ray space of 4D light fields. In *Proc. IEEE CVPR*, pages 1011–1018.

Watson, A. B. 2013. High frame rates and human vision: A view through the window of visibility. *SMPTE Mot. Imag. J*, 122(2), 18–32.

Wendt, G., Faul, F., Ekroll, V., and Mausfeld, R. 2010. Disparity, motion, and color information improve gloss constancy performance. *J. Vis.*, 10(9), 7.

Wetzstein, G., Lanman, D., Hirsch, M., and Raskar, R. 2012. Tensor displays: Compressive light field synthesis using multilayer displays with directional backlighting. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 31(4), 1–11.

Wilcox, L. M. and Stewart, J. A. D. 2003. Determinants of perceived image quality: Ghosting vs. brightness. In *Proc. SPIE*, volume 5006, pages 263–268.

Williem, W. and Kyu Park, I. 2016. Robust light field depth estimation for noisy scene with occlusion. In *Proc. IEEE CVPR*, pages 4396–4404.

Wüller, D. and Gabele, H. 2007. The usage of digital cameras as luminance meters. In *Proc. SPIE*, volume 6502.

Wolff, K., Kim, C., Zimmer, H., Schroers, C., Botsch, M., Sorkine-Hornung, O., and Sorkine-Hornung, A. 2016. Point cloud noise and outlier removal for image-based 3D reconstruction. In *Proc. 3DV*.

Wood, D. N., Azuma, D. I., Aldinger, K., Curless, B., Duchamp, T., Salesin, D. H., and Stuetzle, W. 2000. Surface light fields for 3D photography. In *Proc. ACM SIGGRAPH*, pages 287–296, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.

Woods, A. J. 2012. Crosstalk in stereoscopic displays: A review. *J Electronic Imaging*, 21(4), 040902.

Yeung, S.-K., Tang, C.-K., Brown, M. S., and Kang, S. B. 2011. Matting and compositing of transparent and refractive objects. *ACM Trans. Graph.*, 30(1), 2:1–2:13.

Yu, G., Sapiro, G., and Mallat, S. May 2012. Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity. *IEEE TIP*, 21(5), 2481–2499. ISSN 1057-7149.

Yu, J., McMillan, L., and Gortler, S. 2004. Surface camera (scam) light field rendering. *IJIG*, 4 (04), 605–625.

Yücer, K., Kim, C., Sorkine-Hornung, A., and Sorkine-Hornung, O. October 2016. Depth from gradients in dense light fields for object reconstruction. In *Proc. 3DV*.

Zhang, C. and Chen, T. 2004. A self-reconfigurable camera array. In *Proc. ACM SIGGRAPH 2004 Sketches*, page 151. ACM.

Zhang, F. L., Wang, J., Shechtman, E., Zhou, Z. Y., Shi, J. X., and Hu, S. M. 2016. PlenoPatch: Patch-based plenoptic image manipulation. *IEEE TVCG*, PP(99), 1–1.

Zhang, Z., Wang, L., Guo, B., and Shum, H.-Y. 2002. Feature-based light field morphing. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 21, 457–464.

Zilly, F., Schöberl, M., Schäfer, P., Ziegler, M., Keinert, J., and Foessel, S. January 2013. Light-field acquisition system allowing camera viewpoint and depth of field compositing in post-production. In *Proc. IET*, pages 5.2–5.2(1). Institution of Engineering and Technology.

Zotti, G., Wilkie, A., and Purgathofer, W. 2007. A critical review of the Preetham skylight model. In *Proc. WSCG Short Communications Proceedings I*, pages 23–30.

Zwicker, M., Matusik, W., Durand, F., Pfister, H., and Forlines, C. 2006. Antialiasing for automultiscopic 3D displays. In *Proc. EGSR*, pages 73–82.