Accurate 3D Reconstruction of Dynamic Scenes with Complex Reflectance Properties

Nadejda Roubtsova

Submitted for the Degree of
Doctor of Philosophy
from the
University of Surrey

Centre for Vision, Speech and Signal Processing
Faculty of Engineering and Physical Sciences
University of Surrey
Guildford, Surrey GU2 7XH, U.K.

January 2016

© Nadejda Roubtsova 2016
Summary

Accurate 3D geometry modelling is an essential technology for many practical applications (computer generated imagery, assisted surgery, heritage preservation, automated quality control, robotics etc.). While the existing reconstruction methods mainly operate assuming the simplistic Lambertian model, real scenes, static or dynamic, are characterised by arbitrarily complex \textit{a priori} unknown reflectance properties. The reflectance limitation of the state-of-the-art causes a gap between the practical demand for photometrically arbitrary scene modelling and the constrained applicability scope of existing methods. In response to the gap, this dissertation proposes a solution to the challenging problem of accurate geometric reconstruction of dynamic scenes with arbitrary \textit{a priori} unknown reflectance. This is achieved by introducing a novel approach which generalises Helmholtz Stereopsis (HS) - a niche technique known to be independent of surface reflectance but till now limited to static scenes requiring sequential acquisition of a large number of input views. The undertaken generalisation extends the technique to dynamic scenes by two mutually tailored developments in response to the shortcomings of conventional HS. These developments are 1) a framework to fundamentally improve the geometric reconstruction accuracy from a small set of input images and 2) the design of a novel wavelength-multiplexing-based pipeline for dynamic scene modelling. Together these constitute a novel practical system which, for the first time, enables reconstruction of dynamic scenes with arbitrary surface properties.

To improve the quality of geometric reconstruction by HS, a novel Bayesian formulation of the technique is proposed to replace its sub-optimal maximum likelihood formulation. Further a tailored prior enforcing consistency of per-point depth and normal estimates and related to integrability is developed. The prior purposely exploits the unique ability of HS to characterise the surface by both estimates. The formulation embedded into a coarse-to-fine framework without explicit surface integration achieves unprecedented accuracy and resolution of geometric modelling by HS regardless of reflectance, competitive with what the non-HS state-of-the-art achieves with strictly constrained reflectance.

To generalise HS to dynamic scenes, Colour Helmholtz Stereopsis (CL HS) is proposed which utilises wavelength multiplexing for simultaneous acquisition of the minimal set of input images required for reconstruction. The challenges imposed by wavelength multiplexing in CL HS are addressed using a specially designed calibration consisting of two mutually dependent parts: one infers the photometric properties of the acquisition equipment while the other estimates the reconstructed surface chromaticity spatially and propagates it temporally to accommodate dynamic surface deformation. By integrating the proposed coarse-to-fine Bayesian HS with integrability prior into CL HS, remarkable accuracy and resolution of reconstruction are achieved with the minimal input using just three RGB cameras. Evaluation validates the approach by reconstruction of dynamic scenes with arbitrary \textit{a priori} unknown reflectance, which includes unconstrained spatially varying chromaticity. The reconstructed dynamic sequences exhibit high per-frame geometric accuracy and resolution as well as temporal consistency.

\textbf{Key words:} 3D geometric reconstruction, dynamic scenes, arbitrary BRDF, Helmholtz Stereopsis
Acknowledgements

I would like to take this opportunity to express my profound gratitude to people who have influenced my development as a researcher and supported me on my PhD journey.

First of all, I am deeply grateful to my principal supervisor Dr Jean-Yves Guillemaut who has been, first and foremost, a teacher, but also an ally and a friend in all my scientific endeavours in the past three and a half years. I would like to thank Jean-Yves for taking me on as his PhD student to work on the fascinating topic of 3D reconstruction which I ended up falling in love with. I greatly appreciate him sharing with me the breadth of his knowledge in the field, the theoretical and practical insights, the art of formalisation and rigorous scientific method. Thank you for challenging me to always dig deeper in search for the answer as well as for giving encouragement to help bring down any obstacles in my path. I am also appreciative and very excited about the opportunity I have been given by him to continue the research on complex reflectance at the CVSSP beyond my PhD.

Further, I would also like to thank Prof. Adrian Hilton who in his role as the CVSSP director maintains a friendly and highly stimulating working environment at the centre. The day-to-day functioning of the centre would not be possible without the efforts of the former and present administrative staff (Anna Korzeniowska, Katherine Skinner, Amy Pimperton etc.) and our IT specialists (Dr Helen Cooper, John Field, Bevis King, Borys Jaromin, Dr Bill Christmas etc.) whose door is always open and who greet you with a smile and a helping hand. A special thank you goes to my past and present colleagues (Dr Evren Imre, Alex Freestone, Dr Stuart James, Dr Cemre Zor, Charles Gray, Matt Trumble, Dr Dan Casas and many more!) who have been providing the all-important social element in the everyday office life.

Last but not least, I would like to thank my family for their unfailing support throughout the years. Words cannot express how much I am indebted to my parents, Ella Roubtsova and Serguei Roubtsov, who at an early age inspired me by their own example to pursue an academic career in engineering making a PhD nothing short of a childhood dream. I am also very glad for having been able to share the trials and triumphs of my PhD journey with Mark Brown, my partner in life and a PhD student concurrently with me who has been an irreplaceable source of positivism ever since I met him.
Contents

1 Introduction 3
  1.1 Objectives ........................................ 6
  1.2 Contributions .................................... 7
  1.3 List of Publications ................................ 9
  1.4 Thesis Structure .................................. 10

2 Related Work 13
  2.1 Shape from Silhouette ............................... 14
  2.2 Conventional Stereo ................................ 16
  2.3 Photometric Stereo .................................. 27
  2.4 Shape from Shading .................................. 34
  2.5 Structured light and related algorithms .............. 36
  2.6 3D reconstruction of strongly specular surfaces .... 43
  2.7 Helmholtz Stereopsis ................................. 46
  2.8 Summary and discussion .............................. 51

3 Coarse-to-fine Bayesian Helmholtz Stereopsis 55
  3.1 3D Reconstruction: problem statement and notation . . 58
  3.2 BRDF independent 3D Reconstruction .................. 59
    3.2.1 Helmholtz Stereopsis ......................... 59
    3.2.2 Bayesian Helmholtz Stereopsis ............... 60
  3.3 Prior optimality .................................... 66
  3.4 Implementation ..................................... 68
  3.5 Evaluation ......................................... 71
    3.5.1 Synthetic Data ................................. 71
    3.5.2 Real Data ..................................... 82
  3.6 Conclusion ......................................... 92

4 Colour Helmholtz Stereopsis for Dynamic Scenes with Uniform Chromaticity 95
  4.1 White Light Helmholtz Stereopsis (WL HS) .............. 97
    4.1.1 Photometric Helmholtz camera calibration in WL HS . 100
  4.2 Colour Helmholtz Stereopsis (CL HS) ................... 105
    4.2.1 Acquisition principles and formalisation .......... 106
    4.2.2 Overview of the pipeline ....................... 109
    4.2.3 Photometric Helmholtz camera calibration in CL HS . 109
Contents

4.2.4 Implementation ........................................ 120
4.3 Evaluation ............................................. 124
  4.3.1 Collocation error analysis ......................... 125
  4.3.2 Photometric calibration results .................... 137
  4.3.3 Static scenes ...................................... 138
  4.3.4 Dynamic scenes ................................... 143
4.4 Conclusion ............................................. 147

5 Colour Helmholtz Stereopsis for Dynamic Scenes with Arbitrary Spatially-varying Chromaticity 149
  5.1 Colour Helmholtz Stereopsis with arbitrary chromaticity .... 151
    5.1.1 Formalisation ................................... 152
    5.1.2 The notion of chromaticity ....................... 153
    5.1.3 Overview of the pipeline ......................... 156
  5.2 Surface chromaticity calibration ....................... 157
    5.2.1 Surface chromaticity estimation .................. 157
    5.2.2 Surface chromaticity propagation ................. 163
  5.3 Error analysis ....................................... 166
    5.3.1 Transparency .................................... 168
    5.3.2 Sub-surface scattering/translucency .............. 169
    5.3.3 Reference chromaticity .......................... 174
  5.4 Implementation ...................................... 175
    5.4.1 The pipeline and reference chromaticity .......... 175
    5.4.2 Cross-talk ..................................... 175
  5.5 Evaluation .......................................... 179
    5.5.1 Chromaticity estimation ......................... 180
    5.5.2 Static scenes .................................. 184
    5.5.3 Dynamic scenes ................................ 188
    5.5.4 The challenges ................................ 207
  5.6 Conclusion .......................................... 210

6 Conclusion and Future Work 215
  6.1 Conclusions .......................................... 216
  6.2 Future Work ........................................ 219

Bibliography 221

Appendices 233
Acronyms and Abbreviations

**HS** Helmholtz Stereopsis

**CtF** Coarse-to-fine

**ML** maximum likelihood

**MAP** maximum \textit{a posteriori} probability

**ML HS** Maximum Likelihood Helmholtz Stereopsis

**WL HS** White Light Helmholtz Stereopsis

**CL HS** Colour Helmholtz Stereopsis

**CtF MRF HS** Coarse-to-fine Bayesian Helmholtz Stereopsis solved by MRF optimisation

**CtF ML HS** Coarse-to-fine Maximum Likelihood Helmholtz Stereopsis

**VH** visual hull

**BRDF** Bidirectional Reflectance Distribution Function

**BTDF** Bidirectional Transmission Distribution Function

**BSDF** Bidirectional Scattering Distribution Function

**BSSRDF** Bidirectional Surface Scattering Reflectance Distribution Function

**MRF** Markov Random Field

**SVD** Singular Value Decomposition

**Dprior** Depth-based prior

**Nprior** Normal-based prior

**corr.DNprior** correlation-based depth-normal consistency prior

**dist.DNprior** distance-based depth-normal consistency prior

**PhotoCalib** photometric calibration (of Helmholtz cameras)

**ChromCalib** chromaticity (surface) calibration

**PR** Poisson Surface Reconstruction

**FC** Frankot-Chellappa algorithm

**NoInt** ordering vertices into facets based on known geometric relationships within the reconstruction volume

**HSB** Hue Saturation Brightness

**rms** root-mean-square
Chapter 1

Introduction
Chapter 1. Introduction

The world is a wonderful place blessed with an unfathomable abundance and versatility of form. Equally fascinating is the fact that light can be reflected in many ways depending on the material resulting in a myriad of surface appearances e.g. glossy, specular, opaque etc. The world is also never still but rather in the state of perpetual motion and structural metamorphosis as a result of forces of physics at play. Hence the scenes observed in the everyday life are mostly dynamic, rather than static, and stopping the moment is not an option lest you would lose an interesting part of reality worth capturing encapsulated in the temporal evolution itself. So how can one capture the full intricacy of real geometric form evolving constantly over time regardless surface appearance?

Humankind has always had the inclination to reproduce 3D geometries, first in the form of art and later in the modern times aided by technology. Apart from the purely artistic interest, there are many practical applications of the accurate knowledge of shape. One application is heritage preservation where the earlier man-made copies of the natural world are ironically reproduced for the value in their own right. Film and gaming industries are also interested in 3D reconstruction of real actors and props as these seamlessly integrate with any computer generated content. In mass production, 3D models of parts can be used for automated quality control whereas in the medical domain organ reconstruction may assist diagnosis or guide the doctor during surgery. It is worth highlighting the shared requirement of high accuracy in geometric modelling imposed by all these applications: for instance, consider the imperative photorealism of 3D content in the film/gaming applications which is impossible without the accurate knowledge of surface orientation or the high fidelity to the ground truth both globally and locally of any model used for quality control. Dynamic scenes introduce additional levels of complexity in 3D reconstruction, which results from having only a short time window to collect the necessary information before the scene evolves. The complexity is not simple to overcome and bars the use of some techniques requiring input which is impossible to acquire instantaneously. Due to the high industrial demand and many difficult research problems in the field, 3D reconstruction has been an active research topic from the very early days of computer vision.

Despite its long history, geometric 3D reconstruction is still not universally solved. Human brain recognises 3D geometry effortlessly and impeccably by sampling the reflectance response from visible illumination. Yet the mapping of sampled intensity to geometry is a
difficult problem as a point with a given 3D location and orientation relative to the light source will generate a different response depending on the material’s reflectance properties. In practice, shape and reflectance are fundamentally inseparable in reconstruction based on intensity sampling. This way, reflectance estimation normally requires a good geometric proxy. At the same time, the conventional approach to shape estimation (using which remarkable accuracy has been achieved on tailored surfaces by the best performing geometric 3D reconstruction methods) is to assume an a priori known and/or specific reflectance behaviour. More often than not due to its simplicity and the inherent limitations of the established methods (e.g. Conventional or Photometric Stereo), the assumed photometric behaviour is described by the Lambertian reflectance model, which unfortunately facilitates reconstruction of diffuse surfaces only. If the surface deviates from the Lambertian assumption, Conventional Stereo will fail to establish correspondences between views to compute disparity for depth estimation whereas the normal constraint of Photometric Stereo, traditionally derived from the image formation with the Lambertian reflectance assumption, will no longer be correct. In any attempt to formulate a more realistic normal constraint for Photometric Stereo, there are practical and theoretical difficulties in estimating and parametrising real possibly spatially-varying surface reflectance. The resultant generalisation to a single, specific and rather simplistic reflectance model to constrain shape estimation is a limitation as purely diffuse scenes are not representative of the diversity of the natural and men-made scenes encountered. Evidently there is a supply-and-demand gap in that most scenes are characterised by spatially-varying reflectance of arbitrary complexity and yet there is no method currently in existence capable of accurate reconstruction in the face of such diversity.

The truth is that in order to use reflectance to constrain shape estimation one does not have to model it if generic properties of reflectance are employed. One technique called Helmholtz Stereopsis in particular is notable for exploiting the knowledge of reflectance equality in specific sampling configurations without modelling it. Helmholtz Stereopsis uses the generic physical property of Helmholtz reciprocity to formulate a surface geometry constraint where the exact knowledge of reflectance behaviour is irrelevant. Instead of the absolute knowledge of reflectance different in each case, the universally applicable relative knowledge given a certain sampling geometry is thus used, making Helmholtz Stereopsis
uniquely independent of the reflectance model. Helmholtz Stereopsis is the back-bone of this dissertation because it offers a starting point for the development of a universally applicable 3D geometric reconstruction method for dynamic scenes with arbitrary reflectance.

1.1 Objectives

The global objective of this thesis is accurate high-resolution geometric 3D reconstruction of dynamic scenes with arbitrary \textit{a priori} unknown reflectance properties. It is important to emphasise that the goal is not to model specific classes of photometric complexity and tailor reconstruction to them. Rather the aim is to devise a generically applicable method working well on a wide range of reflectance models without being tailored to any one in particular. Besides the reflectance model freedom, the reconstruction scope is not limited to static scenes. As in most modern 3D research, dynamic scenes are targeted in this work because of the additional challenges they present, e.g. limited acquisition time per frame, variable scene properties every time instant, frame-to-frame reconstruction consistency.

Helmholtz Stereopsis is identified as the most promising option for the foundation of such a system due its inherent reflectance model independence. However, in the present form, the technique does not satisfy the other system specifications in the objective. Firstly, the conventional sequential acquisition procedure of the technique consists in the capture of several pairs of images characterised by the mutually inter-changed camera and light source (i.e. reciprocal pairs). This sampling scenario satisfying Helmholtz reciprocity is achieved mechanically and is not instantaneous preventing application of Helmholtz Stereopsis to dynamic scenes. To exacerbate the limitation, Helmholtz Stereopsis in its current formulation requires a relatively large number of such reciprocal pairs for a reliable reconstruction. Further, even with enough reciprocal pairs, the accuracy and resolution of the technique is often not found sufficient to make it competitive with the modern 3D reconstruction methods.

The thesis aims to develop Helmholtz Stereopsis to satisfy all the requirements specified in the global objective. The task can be broken down into the following specific goals:

1. to improve accuracy and resolution of Helmholtz Stereopsis;

2. to improve robustness of Helmholtz Stereopsis to a reduced number of input images;
1.2 Contributions

The dissertation covers the complete methodology from acquisition to final surface for accurate reconstruction of dynamic scenes with arbitrary a priori unknown reflectance properties. Specifically, the following key contributions have been made in developing the methodology:

1. **Bayesian Helmholtz Stereopsis.** A novel maximum a posteriori probability (MAP) formulation of Helmholtz Stereopsis was proposed to replace the traditional sub-optimal maximum likelihood (ML) one. The MAP formulation consists of a data term based on the Helmholtz reciprocity constraint and a smoothness term exploiting neighbourhood support to improve reconstruction accuracy. The MAP problem is solved by a suitable Markov Random Field (MRF) optimisation method.

2. **Depth-normal consistency prior.** A novel prior tailored to Helmholtz Stereopsis by enforcing consistency between depth and normal estimates was developed for the MAP formulation. The optimality and non-restrictiveness of the prior’s distance-based formulation are shown by deriving its relationship to surface integrability.

3. **Comparative evaluation of priors.** A set of four priors (specifically the depth-based, normal-based, correlation-based depth-normal consistency and the proposed distance-based depth-normal consistency) was evaluated in the context of Bayesian Helmholtz Stereopsis on numerous datasets with versatile geometric and photometric characteristics. The evaluation provides experimental support for the distance-based depth-normal consistency prior.

4. **Coarse-to-fine framework for integration-free reconstruction.** Bayesian Helmholtz Stereopsis was embedded into a coarse-to-fine reconstruction framework to boost reconstruction resolution. The substantial increase in resolution by the coarse-to-fine approach and improved accuracy when using the distance-based depth-normal
consistency prior in Bayesian Helmholtz Stereopsis allow to omit explicit surface integration from the reconstruction pipeline.

5. **Colour Helmholtz Stereopsis.** The novel form of Helmholtz Stereopsis was proposed that exploits wavelength multiplexing and physical camera/light source collocation for instantaneous acquisition of the minimum number of images required for reconstruction. The proposed approach effectively extends Helmholtz Stereopsis to dynamic scenes. Colour Helmholtz Stereopsis utilises coarse-to-fine Bayesian Helmholtz Stereopsis with distance-based depth-normal consistency prior as the reconstruction core without explicit surface integration - a core which is tailored to its drastically reduced number of input images. The method’s pipeline is however also expanded with additional calibration procedures for photometric consistency given multi-spectral illumination. The theory of Colour Helmholtz Stereopsis includes reformulation of the core normal constraint to include photometric characteristics of the set-up and the reconstructed surface.

6. **Photometric calibration procedure.** The procedure calibrating camera sensor sensitivity and light source radiance for traditional Helmholtz Stereopsis using monochromatic illumination is generalised to simultaneous calibration of multiple pairs of RGB cameras and multi-chromatic light sources for Colour Helmholtz Stereopsis.

7. **Chromaticity calibration procedure.** Colour Helmholtz Stereopsis is generalised to dynamic scenes with arbitrary spatially-varying chromaticity by proposing a novel spatio-temporal chromaticity calibration procedure. The procedure seamlessly integrates with the photometric calibration procedure via the use of the same reference together neutralising any signal inconsistency introduced by multi-chromatic sampling.

Contributions 1-4 focus on goals 1 and 2 improving accuracy and resolution with a lower number of input images. The resultant reconstruction framework however at the same time automatically integrates into Colour Helmholtz Stereopsis which is contribution 5 proposed in response to goals 3 and 4. Note that the use of the reconstruction framework in Colour Helmholtz Stereopsis is important as it facilitates accurate geometric modelling.
from just three reciprocal pairs - the maximum number imposed by the employed wavelength multiplexing in the acquisition procedure of Colour Helmholtz Stereopsis. Goals 3 and 4 are reached by means of contributions 5-7.

1.3 List of Publications

The contributions resulted in a number of publications:

**Conference papers:**


**Invited book chapters:**


**Journal papers:**


1.4 Thesis Structure

The thesis is organised as follows:

- **Chapter 2. Related Work**
  
  The chapter presents an overview of 3D geometric reconstruction methods comparing them in terms of accuracy, resolution, reflectance requirements and suitability for dynamic scenes. The comparative analysis provides a justification for the choice of Helmholtz Stereopsis to tackle the objectives of this work.

- **Chapter 3. Coarse-to-fine Bayesian Helmholtz Stereopsis (contributions 1-4)**
  
  The chapter develops and evaluates the proposed Bayesian formulation of Helmholtz Stereopsis with the novel tailored distance-based depth-normal consistency prior related to surface integrability. A pipeline without explicit surface integration involving a coarse-to-fine approach is introduced.

- **Chapter 4. Colour Helmholtz Stereopsis for Dynamic Scenes with Uniform Chromaticity (contributions 5,6)**
  
  The chapter proposes Colour Helmholtz Stereopsis that for the first time allows dynamic scene reconstruction with arbitrary reflectance. The discussed key aspects of the method include a novel multi-spectral acquisition set-up, the generalised Helmholtz reciprocity constraint formulation for multi-spectral illumination, the tailored reconstruction core proposed in Chapter 3 and the photometric calibration for inter-channel signal consistency in the set-up. In this chapter the developed calibration procedure for the multi-spectral photometric set-up with respect to an arbitrary reference chromaticity sets Colour Helmholtz Stereopsis up for reconstruction of any surface with uniform chromaticity corresponding to the reference.

- **Chapter 5. Colour Helmholtz Stereopsis for Dynamic Scenes with Arbitrary Spatially-varying Chromaticity (contributions 5,7)**
  
  The formulation of Colour Helmholtz Stereopsis is further generalised to allow for spatially varying arbitrary (non-reference) surface chromaticity. To this end, a novel
spatio-temporal chromaticity calibration procedure for dynamic scenes is proposed and integrated into the pipeline.

**Chapter 6. Conclusion and Future Work**

The chapter draws conclusions on the content of the thesis as a whole reflecting on how well the objectives of the work have been met. The strengths as well as the existing limitations of the work are identified and directions for future work are proposed.

**Appendices**

Some additional derivations and the exhaustive set of results relating to Chapter 3 are provided at the end of the thesis.
Chapter 2

Related Work
3D geometry reconstruction has been an active research topic in computer vision and robotics from mid-1970s. The field can boast a wide range of methodologies with varying degrees of complexity of acquisition and processing as well as characteristic strengths and limitations. The aim of this chapter is to provide a literature survey on 3D reconstruction with methodology classification charting out the field for easier navigation. The focus of the survey is on the techniques assuming full geometric calibration of the cameras, which includes both their intrinsic (i.e. focal length, principle point, distortion) and extrinsic (i.e. 3D position and orientation) parameters. As the topic of this dissertation is accurate dynamic reconstruction with arbitrary reflectance complexity the presented prior art approaches will be systematically assessed in terms of suitability for this goal. The systematic analysis provides solid grounds for the choice to develop the little known niche method of Helmholtz Stereopsis as the most novel and fundamentally most promising approach for the target application.

The vast majority of 3D reconstruction methods are intensity-based meaning that the constraining information is obtained by measuring pixel values on acquired imagery although other cues, such as the knowledge of acquisition geometry, can be exploited in conjunction with the measurements. These techniques, forming the most dominant category in 3D reconstruction with validated performance, are the focus of the literature survey. Intensity-based methods are versatile and can be further sub-divided into many classes differing in key aspects such as acquisition principles and set-up, underlying reconstruction assumptions, surface characterisation etc. Before delving into these methods however, let us briefly discuss Shape from Silhouette, which exemplifies a category of geometric reconstruction techniques making no use of intensity information.

2.1 Shape from Silhouette

The methods in this category, which is one of the oldest, are purely geometric drawing information from the knowledge of the calibrated acquisition set-up and performing no intensity sampling on the acquired imagery beyond the binary check of whether the point is in or out of the region of interest. The region of interest is determined by multiple silhouette masks, which are typically user-defined in natural environments but, in the context of chroma keying, can also be automatically obtained. The methods are collectively known
2.1. Shape from Silhouette

as Shape from Silhouette. The original Shape from Silhouette method was introduced in [1]. The output of Shape from Silhouette is the visual hull [2] - a rough 3D outline of the object computed by intersecting visual cones from multiple views. The visual hull is considered a coarse representation in itself and is used as the starting point by many complex reconstruction algorithms.

Traditionally, there have been two classes of Shape from Silhouette methods: the volumetric and the surface-based. The methods in the volumetric class, e.g. [3], [4], discretise the space into voxels and carve the surface boundary by selecting the voxels projecting within all the silhouette masks. The surface-based approaches, e.g. [5], [6], work on the points on the interface between the interior and exterior of the visual hull only. In surface-based Shape from Silhouette, reconstruction is performed by integrating visual cone surfaces corresponding to occluding contours defined by the silhouette masks. The volumetric methods typically produce coarse results due to the finite discretisation resolution of the 3D space and are computationally expensive. The surface-based methods, such as the recent [6], are more precise but can lack robustness due to visual cone boundaries from which contour generators are derived often being ill-defined. Hybrid approaches combining the best aspects of volumetric and surface-based classes are possible. Boyer and Franco in [7] perform discretisation of space, typical of volumetric methods, but do so only on the surface of the visual hull, instead of uniformly throughout the volume. In short, a set of surface points of the visual hull is computed by triangulation, computing the depths of the intersection points of the epipolar line with the occluding contours for each silhouette mask. Delaunay triangulation of the surface points produces the convex hull, which can then be refined by checking centroid re-projection of computed tetrahedra into the silhouette masks.

Automatic silhouette extraction in natural environments remains an open problem. Simple background subtraction often fails due to, for instance, noise or inherent ambiguities in the scene. It has also been pointed out that monocular silhouette definition may not be consistent between camera views. Recently, in response to the issues, there has been research on computing visual hull probabilistically, by Bayesian inference instead of pre-defined silhouette masks. In [8], a Bayesian formulation of the problem of voxel occupancy likelihood based on noisy image data is solved. Voxel occupancy is related to the variables of observed background and foreground+background intensity, any prior knowledge of the
acquisition system or the scene and the hidden variable of binary silhouettes. In this context there are thus no pre-defined silhouette masks. The strength of the approach is the ability to use the entire scope of available information with causality explicitly modelled which allows to avoid irrevocable binarisation \textit{a priori} based on incomplete information. Another notable example of a probabilistic visual hull approach in [9] is different in its use of class-specific human silhouette prior, instead of an arbitrary one.

Despite all the recent advances mentioned, Shape from Silhouette techniques are still limited by their fundamental inability to reconstruct concavities and the severe dependence of the performance on the number of views available. The latter limitation makes Shape from Silhouette inapplicable for dynamic scene reconstruction to the desired degree of accuracy and resolution. Traditional Shape from Silhouette techniques with binary masks are by definition reflectance model independent due to the fact that intensity is not sampled. The newer probabilistic visual hull variants do sample intensity and hence will also be affected by any inconsistency of surface point appearance due to non-Lambertian reflectance.

2.2 Conventional Stereo

The Principle. Projective geometry of image acquisition described by the geometric calibration of the camera defines how points along the projection ray map onto an image pixel. Since all points on the ray map onto the same pixel, one can speak of depth ambiguity in the process of back-projection of an image pixel to the 3D world. Conventional Stereo solves the ambiguity by considering a pair of images obtained by geometrically calibrated cameras with overlapping views. The so-called epipolar geometry in a Conventional Stereo configuration consisting of a pair of cameras ensures that a projection ray, perceived as a point in the reference image $I_r$ of one camera, will be observed as a line, i.e. a set of distinguishable points, in the matching image $I_m$ of the other camera. This is the basic principle allowing to establish which point on the projection ray to $I_r$ is actually sampled. Conventional Stereo performs triangulation in the described configuration computing disparity between the two images. Disparity is the relative displacement of corresponding feature points along the epipolar lines after rectification. An epipolar line corresponding to a given projection ray is defined as the line between the projection of the ray onto the image plane of one camera and the camera’s epipole. The epipole is in turn defined as
the projection of the centre of the second camera in the epipolar configuration onto the image plane of the first camera. Epipolar geometry of the configuration is characterised by how epipolar elements of the pair of cameras are positioned relative to one another. The process of rectification aligns the epipolar lines of the pair of cameras based on their known geometric calibration characteristics. The procedure essentially amounts to moving both epipoles to infinity.

The first stage of Conventional Stereo is feature point matching which involves the search for correspondences along the rectified epipolar lines. To this end, the stage needs the presence of sufficient texture and a uniformly Lambertian reflectance model of the surface to guarantee the same feature point appearance regardless of the viewpoint. Mathematically, image points \((u, v)\) and \((u', v')\) in \(I_r\) and \(I_m\) are matched. When \(I_r\) and \(I_m\) are rectified \((v = v')\), one can write \(u' = u + d(u, v)\) where \(d\) as a function of pixel location is the disparity.

The degree of appearance similarity between pixels \((u, v)\) and \((u', v')\) is the matching cost of the assignment of \(d(u, v)\) to point \((u, v)\) of \(I_r\).

The second stage of Conventional Stereo is triangulation to compute depth from disparity. By triangulation, disparity \(d\) is inversely proportional to depth \(z\). Depth \(z\) corresponding to projection \((u, v)\) in the reference image \(I_r\) is thus computed by:

\[
z = \frac{bf}{d} \tag{2.1}
\]

where \(b\) is the baseline between the cameras and \(f\) is the focal length (equalised between the cameras as a result of rectification).

Since Conventional Stereo is such an old class of methods, the amount of literature on the topic is vast. Broadly, the techniques in the class can be divided into 2-frame and multiview respectively corresponding to 2.5D and full 3D reconstruction.

**2-frame.** In response to lacking transparency, Scharstein et al. [10] have produced a well thought-through taxonomy of approaches falling under the umbrella of 2-frame stereo. The classification rules they propose facilitate a more systematic comparison of 2-frame stereo variants allowing one to scrutinise individual system components independently. The authors also propose an evaluation procedure with categorised datasets and benchmarks to quantify comparisons. This Middlebury Stereo evaluation procedure [11] has gained wide acceptance in the stereo vision community. An up-to-date table of the top-performing al-
Scharstein et al. identify a number of classification criteria. Firstly, with regards to feature matching, the algorithms can be either sparse or dense. For more practical dense algorithms, the taxonomy recognises four stages presented in Table 2.1, each with a variety of possible realisations. The set of choices made for each component defines the type of the whole algorithm. The stage-level choice of approach is however not entirely independent of the other stages. In the presented classification system, choosing mutually tailored stages makes the whole pipeline classify as either a local, global or an in-between iterative 2-frame stereo algorithm. Depending on the class, the role of any given stage may become more or less prominent.

2-frame Local. Local approaches are window-based meaning that pixel disparity is dependent on a set of its closest neighbours. In these algorithms, piecewise smoothness is enforced through cost aggregation. Easily parallelisable and portable onto the GPU, these algorithms are faster compared to the global alternatives. Recent work presenting local Conventional Stereo [12] supports the earlier conjectures that scan-line optimisers are capable of delivering top reconstruction accuracy, rivalling that of global optimisers, provided appropriate modifications are made to the original naïve one-directional implementation. Mei et al. in [12] conclude that the GPU-targeted implementation results in a speed-up of 140 times: all datasets take under 0.1 s. The algorithm produces high-ranking results but they are not equally good on all Middlebury datasets. The real-time performance inspired testing on videos. Unfortunately, these results are not as convincing because of noise, rectification errors and varying illumination causing problems for cost computation and support region construction.

2-frame global. Global stereo methods, on the other hand, formulate a classical MAP optimisation problem in which the smoothness assumption is made explicit through one of its terms. These techniques show consistent accuracy and robustness by minimising the total penalty cost over the disparity map [13]. MRF techniques, particularly graph cuts, have proven their merit in solving such MAP problems both in 3D reconstruction and other computer vision problems. As explained in [14], the wide applicability is due to the fact that the majority of computer vision tasks, both high and low level, such as segmentation,
<table>
<thead>
<tr>
<th>Nr.</th>
<th>Stage</th>
<th>Aim</th>
<th>Design space (e.g.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>matching cost computation</td>
<td>to establish the degree of certainty in assigning a given disparity to a pixel location (equivalently, the certainty of a feature match)</td>
<td>1. squared intensity difference/absolute intensity difference (equivalent to the mean-squared-error and mean absolute difference); 2. cross-correlation; 3. binary matching and 4. gradient-based measures.</td>
</tr>
<tr>
<td>2</td>
<td>matching cost aggregation</td>
<td>to introduce continuity between individually computed pixel costs (effectively, averaging over some support region); More important for local algorithms while global algorithms typically work with un-aggregated costs shifting complexity to the optimisation stage; Aggregation for global algorithms may provide initial estimates for iterative procedures or identify sets of high confidence matches.</td>
<td>1. Square; 2. Gaussian; 3. with/without adaptive behaviour and/or truncation.</td>
</tr>
<tr>
<td>3</td>
<td>disparity computation and optimisation</td>
<td>to generate a disparity map conforming to the piecewise smooth assumption (the single most important stage for global methods).</td>
<td>1. winner-takes-all (trivial, local methods); 2. graph cuts, max flow and other MRF techniques; 3. simulated annealing; 4. dynamic programming; 5. scanline optimisation (fast, reaching global minimum for independent scalines in polynomial time even if the cost function is NP-hard).</td>
</tr>
<tr>
<td>4</td>
<td>disparity refinement</td>
<td>to eliminate discretisation effects for sub-pixel accuracy since unless continuous optimisation methods are chosen in stage 3 (e.g. splines) the disparity map will be spatially discrete.</td>
<td>1. iterative gradient descent; 2. curve fitting.</td>
</tr>
</tbody>
</table>

Table 2.1: Four stages of a typical conventional 2-frame stereo and the design space of possible realisations of each stage from [10]. The set of choices made for the four stages determine the place of each particular Conventional Stereo algorithm in the classification system of Scharstein et al.
image restoration, object matching or disparity computation, can be modelled as a labelling problem. The optimal solution to such a problem is one maximising the *a posteriori* probability estimate. The posterior probability is the likelihood of a parameter assuming a certain value given the observation. The problem is typically translated into the equivalent one of posterior energy minimisation where the total energy is a linear combination of likelihood and prior energy terms. The likelihood is related to the noise model of the observation representing its quality when viewed independently of the other observations. The prior term expresses the set constraints or the prior knowledge of the particular problem e.g. (piece-wise) smoothness of real world surfaces. In the prior, one usually encapsulates the interaction between neighbouring observations.

In global conventional 2-frame stereo, disparity assignment is formulated as an MRF labelling problem i.e. as an energy minimisation of a Bayesian cost function. Disparity is the label assigned per pixel of the reference image. Each disparity option per pixel is evaluated based on an energy function combining likelihood and prior terms. The likelihood term in conventional 2-frame stereo is normally quantified by some form of intensity comparison (e.g. the squared/absolute intensity difference, cross-correlation, binary matching, gradient-based measures) between the two pixels indicated as corresponding by a given disparity assignment. The penalty for intensity discrepancy depends on the assumed noise threshold or, generally speaking, on the adopted noise model. The prior term in the disparity map optimisation typically describes smoothness, penalising abrupt changes in depth from one pixel to another (since the system is Markovian, only neighbouring pixels are compared). Despite being the best performing in terms of accuracy [10], global techniques are notoriously slow. Dynamic programming and scan-line optimisation techniques would afford a speed up at the expense of accuracy as they solve a simpler problem optimising only horizontally.

MRF energy minimisation algorithms applicable to Conventional Stereo are numerous and mature. A comparative study by Szeliski *et al.* [15] aims to find the optimal ones for the specific class of energy minimisation problems with smoothness-based priors characteristic of typical computer vision applications. The results of the study have shown *graph cuts-Expansion* [16] and *Sequential Tree Re-Weighted message passing* [17], [18] to be the two best performing methods, coming to within 1% of the global minimum. Expansion is also
2.2. Conventional Stereo

the absolute winner in terms of run-time. Furthermore, there does not ever seem to be any reason to use graph cuts-Swap over graph cuts-Expansion [16]. The limitation of graph cuts in general is the requirement for the smoothness term of the minimised cost function to be regular (regularity is defined by Kolmogorov and Zabih [19]). If the smoothness term is not regular, the cost energy is said to be not graph-representable and the energy minimisation problem becomes intractable (NP-hard) and impossible to solve by graph cuts [19]. There has been work on extensions to optimisation with highly descriptive second-order priors [20] previously unusable in global optimisation algorithms. Another interesting recent direction of research in global optimisation for Conventional Stereo has been the use of 3-dimensional labels i.e. depths + surface orientation instead of just the depths used traditionally [21].

2-frame iterative. Finally, iterative algorithms are hierarchical (coarse-to-fine) in nature meaning that there is no global energy minimisation function. Instead progressively higher levels of precision are computed, with the coarser level results being used to constrain more local searches. The main weakness of iterative approaches is their dependence on initialisation without which they can get stuck in a local minimum [22]. The iterative algorithms are highly suitable as a refinement step [23] after the global shape has been established by an alternative method.

Conventional 2-frame stereo is a highly developed mature technique. As a 2-frame approach, the technique generates a 2.5D reconstruction only. Multiview techniques discussed in the following section extend the method to full 3D by expanding input data beyond a single pair of images and tackling the task of merging the partial reconstructions into a single model. Since the 2-frame Conventional Stereo lies in the core of multiview the basic classification classes of local, global and iterative are equally suitable for multiview classification. In the next section extensions to the system and some additional classification aspects typical of multiview specifically are discussed.

Multiview. A logical extension of 2-frame Conventional Stereo is reconstruction of full 3D models of objects using a set of more than two images taken from multiple viewpoints by having to tackle data fusion challenges. A taxonomy of multiview algorithms in [24] presents a classification of the multiview Conventional Stereo techniques. The work comprises a
clear classification system, an evaluation methodology featuring the measures of accuracy and completeness, calibrated multiview datasets and an evaluation of characteristic top performing algorithms. This evaluation methodology is the guideline for the Middlebury multiview algorithm ranking [25].

**Classification and examples.** In their work, Seitz et al. limit the scope to algorithms reconstructing dense object models from calibrated views, excluding binocular (see the previous section on conventional 2-frame stereo), trinocular, multi-baseline and structure-from-motion approaches. According to Seitz et al. [24], the vast behavioural variation of multiview algorithms arises from the choices made for the six fundamental aspects of the system presented in Table 2.2, specifically 1. scene representation, 2. photo-consistency measure, 3. visibility model, 4. shape prior, 5. reconstruction method and 6. initialisation requirements.

Most importantly, for the reconstruction method (number 5 in Table 2.2) Seitz et al. identify: 1. cost function minimisation (global or iterative); 3. depth map integration enforcing consistency and 4. surface fitting to extracted features. In other categories, there are classes such as image-space vs. scene-space photo-consistency or voxel-based vs. level-set-based scene representation. Traversal of the voxel grid scene representation for reconstruction appears in early iterative multiview voxel colouring [26]/voxel carving [27]. One Middlebury top performer [28] starts off with a feature-based approach followed by transformation of a set of patches into a polygonal mesh by an iterative joint optimisation of smoothness, photometric and rim consistency functions. This method by Furukawa and Ponce thus consists of two separable reconstruction phases. In the first phase, features, extracted by Harris-corner and Difference-of-Gaussians operators, lie in the centre of patches that are matched between calibrated views using robust, though costly, region-based photo-consistency. After expansion for a denser patch-based surface representation and filtering to remove erroneous matches, a *surfel model* of the object consisting of a dense set of oriented patches results. By computing the convex hull of the reconstructed points, Furukawa and Ponce automatically obtain the bounding volume unlike other prior art approaches where it needs to be initialised. This bounding volume is the starting point if it is desired to convert the recovered collection of patches into a surface mesh. The transformation of a series of patches into a surface mesh (i.e. polygonal surface reconstruction) is done by repeatedly
### Table 2.2: The six key aspects of a multiview algorithm determining its place in the classification system of Seitz et al. [24].

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Design space (e.g.)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scene representation</td>
<td>1. voxels; 2. level-sets; 3. polygonal meshes or depth maps;</td>
<td>Voxels i.e. a discrete occupancy function and level-sets i.e. the distance to the nearest surface - regularly sampled volumetric representations, a popular choice for simplicity, uniformity and the ability to approximate any surface. Polygonal meshes - valued for their storage and rendering efficiency. Depth maps - 2D representation avoiding conversion to the 3D domain.</td>
</tr>
<tr>
<td>Photo-consistency measure</td>
<td>1. scene-space; 2. image-space;</td>
<td>Scene-space approaches (favour small surfaces) - projection of a unit of scene representation onto corresponding multiview images. Image-space (biased towards frequently occurring elements/those occupying vast image areas) - photometric prediction error when an image is warped from one view to the next using the current estimate of scene geometry.</td>
</tr>
<tr>
<td>Visibility model</td>
<td>1. geometric; 2 quasi-geometric 3. outlier-based.</td>
<td>i.e. occlusion handling; geometric - explicitly inferring image sets where individual scene structures are visible; quasi-geometric - infer visibility relationships between cameras; outlier-based - treating occlusions as outliers.</td>
</tr>
<tr>
<td>Shape prior</td>
<td>1. minimal surface; 2. maximal surface; 3. reference shape (a sphere, plane etc.); 4. smoothness.</td>
<td></td>
</tr>
<tr>
<td>Reconstruction method</td>
<td>1. global cost-function-based; 2 iterative cost-function based; 3. depth map consistency based; 4. feature-based;</td>
<td>global - MRF optimisation (graph cuts, max flow); iterative - gradual coarse-to-fine surface evolution; depth map consistency - computes and integrates depth maps; 4. feature-based - extracts features to which the surface is subsequently fitted.</td>
</tr>
<tr>
<td>Initialisation requirements</td>
<td>1. bounding box/volume; 2 background/foreground segmentation; 3 visual hull; 4 realistic depth range</td>
<td></td>
</tr>
</tbody>
</table>
moving each vertex on the bounding volume of the initial phase according to the forces of
1. surface smoothness (i.e. shape prior), 2. photometric consistency and 3. rim consistency
for datasets where an accurate silhouette is available. In the terminology of Seitz et al.,
this second (optional but normally desirable) phase of conversion is of the iterative type
which is here once again used for refinement. Furukawa and Ponce in [28] cannot handle
dynamic scenes but they have demonstrated reconstruction with successful elimination of
transitory occluders e.g. water running down the reconstructed steps and people occluding
various parts of a building of interest are both successfully filtered out as outliers.

Reflectance model. Recently, there has also been work [29] [30] [31] striving to allevi-
ate the inherent restriction to Lambertian reflectance shared by all Conventional Stereo
approaches due to the fundamental reliance on photo-consistency. Jin and colleagues
in [29] [30] explicitly model non-Lambertian behaviour of scenes and propose a novel
model-to-image discrepancy measure for non-Lambertian surfaces based on the more generic
Ward’s reflectance model and formulated via the radiance tensor containing multiview ob-
servations of reflectance per surface patch. The work is interesting as multiview explicitly
targeting non-Lambertian surfaces in its formulation is very scarce but the approach re-
mains constrained by the assumption of a given reflectance model, albeit a more generic one.
In another very recent multiview work seeking to get rid of reflectance constraints [31] a
different approach is taken. Instead of explicit reflectance modelling, the approach aims for
joint global estimation of shape and reflectance properties from a set of multiview images.
Although the idea of joint inference of unknown parameters (i.e. shape and reflectance)
based on observable characteristics (i.e. intensity and the illumination conditions) is plausi-
ble, the two estimates will remain inherently linked and the shape cannot be expected to be
more accurate than the reflectance estimate whose full complexity is difficult to model with
a finite number of images. Although conceptually interesting, non-Lambertian multiview
methods comparatively do not deliver particularly accurate or high resolution results.

Dynamic scene reconstruction. Recent multiview work [32–36] presented systems for
non-rigid dynamic scene reconstruction. Furukawa and Ponce in [32] propose a dual-process
tracking procedure tracing the deformation of an instantaneously acquired mesh. One of
the processes tracks the local rigid deformation of each vertex while the other takes care of
the global non-rigid deformation of the model as a whole. The algorithm was later further
2.2. Conventional Stereo

improved in [33] to handle more extreme tangential non-rigid deformation specifically those involved in face capture. A more recent facial capture system using multiview [34] propagates geometry throughout the sequence using the automatically detected anchor frames that are similar to the chosen reference facial expression. Such tracking is used as a tool for refinement of per-frame reconstructions for temporal consistency. Unlike the previous methods, in [36] full-human-body reconstruction down to the resolution of garment cloth deformation is tackled. A pipeline is introduced that fuses silhouette, feature and stereo cues. The strength of the pipeline is in the progressive coarse-to-fine reconstruction: the visual hull constrains the scene, edge-based features establish initial correspondences for sparse reconstruction and finally through a global optimisation of colour/intensity costs dense reconstruction refines the shape. Guillemaut and Hilton [35] present multiview reconstruction from low-resolution sport videos for freeview generation using wide baseline cameras calibrated on the fly. For robustness against calibration inaccuracies a view-dependent approach is adopted. A loose view dependency is introduced through a novel multiview consistency prior constraining layered depth estimation of a view using depth estimates of neighbouring cameras. The key feature of the work is joint segmentation and reconstruction: i.e. separation of the individual foreground layers from the background together with depth recovery in the foreground. The joint approach avoids error propagation between stages and allows simultaneous utilisation of shared cues. The work performs optimisation based on the metrics of colour, contrast, photo-consistency, smoothness and segmentation/depth consistency between views and has shown competitive results on static Middlebury datasets as well.

Large-scale (real-time) modelling. As the majority of Conventional Stereo algorithms in general, a lot of multiview algorithms are based on intensity matching. One notable exception is the recent photo-tourism work [37,38] on sparse outdoor reconstruction based on large quantities of unrelated photos from Internet repositories. These impressive photo-tourism algorithms match views based on suitable interest points [39], e.g. invariant SIFT features [40], not directly described via photo-consistency. The issue remains however that such interest points facilitate only a sparse reconstruction which limits their applicability as a consistency measure in Conventional Stereo. The real demand is actually for dense 3D reconstruction of large-scale (e.g. outdoor) environments and hence it has been a much
researched direction in the multiview community recently. In [41], a monocular motion stereo approach is proposed using a fish-eye camera characterised by its wide field of view. The camera poses in this type of approaches are not pre-calibrated but derived by motion tracking. There is clearly a similarity in the paradigm of the method to other popular depth map fusion techniques discussed later on (e.g. KinectFusion) which however differ in their use of a depth sensor for raw depth data acquisition. Depth maps by stereo matching are said to be more prone to outliers than those obtained with a sensor. Hence filtering procedures are proposed trading completeness for accuracy. As with all accumulative approaches gradually building up the model of the environment, completeness can still be improved using subsequent measurements under user guidance for view planning. Stereo matching in [41] is performed using the plane sweep stereo approach [42] run on the GPU of a hand-held mobile device, which enables real-time performance of the technique. The plane sweep stereo from [42] is a multi-resolution approach using the standard sum-of-square-differences similarity metric implementable on the GPU. The multi-resolution aspect of the approach provides robustness in the areas of depth discontinuity and low texture. The stereo matching technique is also used by an earlier work on urban 3D reconstruction [43] employing many of the same principles as [41]. Unlike earlier photo-tourism algorithms, the models presented in [41] are (piecewise) dense and with a greater resolution. The method’s limitations include the inability to deal with textureless surfaces and loop closures in the motion trajectories.

**Assessment of Conventional Stereo techniques.** The breadth of research on mature Conventional Stereo has been seen to produce impressive results in static and dynamic scene reconstruction. There are however limitations on the type of scene that can be reconstructed with Conventional Stereo. Conventional Stereo methods inherently make the Lambertian reflectance model assumption to guarantee photo-consistency between views without which feature matching is not possible. Any violations of the Lambertian assumption can be handled to some extent in complex optimisation schemes but never solved for completely making the assumption a major limiting factor of intensity-feature-based Conventional Stereo methods. The alternative point-of-interest-based Conventional Stereo methods are more robust to inter-view intensity variation but unfortunately are at the same time limited to sparse reconstruction only. These methods also rely on the presence of suffi-
cient texture, more so than the intensity-feature-based ones. Attempts at non-Lambertian Conventional Stereo have not been compelling enough so far to become mainstream. Further, disparity-based reconstruction of Conventional Stereo characterises the surface by a set of depth estimates only. Although a robust descriptor of the global shape, depth-based reconstructions tend to be of a lower structural resolution than those characterising surface points by normals, lest an extremely spatially fine surface sampling is undertaken at a burgeoning computational cost. This limitation of exclusive characterisation by depth has led to the development of hybrid techniques (discussed further on) combining Conventional Stereo with other methods to be able to simultaneously generate normal information in addition to depth.

2.3 Photometric Stereo

The principle. Photometric Stereo, first introduced by Woodham in the 1980s [44], [45] measures a surface point’s response to varying illumination at a constant viewpoint meaning that, unlike Conventional Stereo, no feature point matching is required. The technique is called photometric because, instead of relative feature displacements, irradiance values at a single image position are measured. A dense field of estimated surface normals is produced at the output. The key principle of Photometric Stereo can be made intuitively clear [44], [45] by defining reflectance field \( R(p, q) \) as a function of local surface orientation \([p, q, -1]\) where \( p = \frac{\partial z}{\partial x}, q = \frac{\partial z}{\partial y} \) and the negative \( z \) is aligned with the viewing direction. In a simplified model, image intensity \( I = I(u, v) \) is equal to \( R(p, q) \). At least a triplet of intensities \([I_1, I_2, I_3]\) of pixel \((u, v)\) is measured corresponding to the three illumination directions in matrix \( L_{3 \times 3} = [L_1, L_2, L_3] \). Each illumination direction \( L_n \) where \( n = [1, 2, 3] \) is a 3-vector. Using the a priori known \( R(p, q) \), iso-brightness contours can be constructed: \( R_1(p, q) = I_1, R_2(p, q) = I_2, R_3(p, q) = I_3 \) in the \( pq \) plane. The \((p, q)\) coordinates where all three (or more) iso-brightness contours intersect in the \( pq \) plane are the sought surface orientation components. In other words, the sought surface orientation is the one best reconciling intensity prediction \( R(p, q) \) and the measurement matrix \( I_{3 \times 1} = [I_1, I_2, I_3]^\top \).

In practice, in the case of the commonly assumed Lambertian reflectance model and taking the generic case of \( n \) observations under varying illumination conditions, the normal scaled by albedo \( \rho \) at the sample surface point is computed by inversion of the illumination
direction matrix \( L_{3\times n} \) given the observed intensity measurement matrix \( I_{n\times 1} \) [46]:

\[
L_{3\times n}^{-1}I_{n\times 1} = \rho n
\]  

(2.2)

The orientations of the estimated vectors \( \rho n \) is the sought normal field. In addition, an albedo map of the surface is generated from the set of magnitudes \( \rho \) of all visible surface points. Mathematically, the general formulation of albedo is [46]:

\[
\int_{-\infty}^{\infty} E(\lambda) R(\lambda) S(\lambda) d\lambda
\]  

(2.3)

where \( E(\lambda) \), \( R(\lambda) \) and \( S(\lambda) \) define respectively the spectral characteristics of the light source, reflectance of the surface and spectral sensitivity of the camera sensor. In the case of white light illumination and capture by a camera with a spectrally flat sensor, albedo is purely \( \int_{-\infty}^{\infty} R(\lambda) d\lambda \) i.e. the surface appearance due to its reflective properties independent of surface orientation which is in turn encoded in the normal estimate.

Unlike Conventional Stereo, Photometric Stereo is not inherently limited in its applicability to surfaces with Lambertian reflectance. Equation 2.2 illustrates the key idea of Photometric Stereo taking the Lambertian reflectance model encapsulated in a single albedo constant \( \rho \) per surface point as an example because of its simplicity: the constant is independent of the normal orientation. At the peril of a more complex formulation any parametric reflectance model can be used. However, the choice to simplify to the Lambertian model is made all too often in Photometric Stereo research owing to the known difficulty and tediousness of reflectance model estimation in practice.

**Reflectance model.** The accuracy of Photometric Stereo is known to suffer from inaccurate knowledge of the reflectance model. Impressive research efforts have been made towards both error correction and prevention in tackling this problem. For example, to mitigate the corrupting effect of non-Lambertian behaviour amongst other things, Wu *et al.* [47] propose a highly effective MRF formulation for Photometric Stereo with prior-based regularisation of normals. Their impressive results show a remarkable robustness to noisy input, complex geometries, specularities, shadows and even transparencies. On the front of reflectance model error prevention, one could mention the interesting work [48] on arbitrary BRDF Photometric Stereo by example. The idea is to find relationships between
2.3. Photometric Stereo

surface normals and observed reflectance behaviour by sampling reference objects of known geometry under different illumination. The test object’s material is considered a linear combination of reference materials. With the assumption not being fundamental, the work however cannot claim to have universally solved reflectance model dependence. Further, self-calibration of the reflectance model during reconstruction, instead of as a separate procedure, is a desirable property. Research has been done on how to obtain the unknown reflectance information for Photometric Stereo at reconstruction time. In [49], together with a self-calibration procedure of camera pose/light direction, the authors address reconstruction of multi-albedo Lambertian surfaces by Photometric Stereo. In this work specularities are treated not as violations of the Lambertian model but rather as albedo variation which is effectively calibrated during reconstruction. The method has been shown capable of handling specular highlights without adopting a more complex reflectance model in the formulation of the Photometric Stereo by optimising for surface albedo explicitly, together with the surface normal, for every vertex over the set of images in which it is visible. The estimation of a highlight as variable albedo is not a principled approach as it is not a surface property and is in fact dependent on the direction of illumination and the normal. However, due to a degree of similarity in illumination conditions in the set of images used for vertex reconstruction, the highlight may indeed be stable enough to be treated as albedo. The method has been shown capable of reconstructing static scenes that feature isolated specular highlights (e.g. porcelain figurines) using 36 images acquired with a turn-table and an uncalibrated camera/light source set-up. A principally different reflectance self-calibration approach was proposed by Vogiatzis and Hernández [50] who address the difficulty of non-Lambertian reflectance in Photometric Stereo by fitting a Phong model [51] based on the outliers from the Lambertian model estimation procedure. For problem tractability, the assumptions of monochromaticity and (spatially and spectrally) constant specular model parameters are made. In the formulation with the more complex Phong reflectance model, the model’s invertibility issues are tackled using the mentioned assumptions. The formulation is innovatively non-Lambertian, but the scene reflectance model allowed by the method is far from arbitrary but in fact limited by the many assumptions made (Phong model, monochromaticity, constant specular components parameters etc.) in pursuit of tractability and well-posedness of the mathematical formulation of the
Photometric Stereo problem. Further there has also been work on exploiting isotropy (i.e. rotation invariance about the normal) as a property shared by a majority of real-life reflectance models [52] [53]. The assumption of isotropy in Photometric Stereo allows the construction of the so-called iso-depth contours where all points are equidistant from the image plane. In [53], impressively accurate geometries of non-Lambertian surfaces are reconstructed by propagation of sparse surface points obtained by Structured from Motion along the constructed iso-depth contours. Any work based on a common property of a large class of reflectance models is interesting as a step towards reflectance model generalisation. However, isotropy is not a generic property of the Bidirectional Reflectance Distribution Function (BRDF), unlike energy conversation and reciprocity, meaning that the method is still not universally applicable. Relying on isotropic reflectance to build continuous iso-depth contours, these methods will also be sensitive to cast shadows and inter-reflections.

Common BRDF symmetries induced by both wide-spread isotropy and generic reciprocity are similarly considered in [54] in the context of Photometric Stereo for reconstruction of realistically non-Lambertian scenes. In Section 2.7 later on in this chapter an entirely different (non-Photometric Stereo) reconstruction approach called Helmholtz Stereopsis based exclusively on the generic reciprocity property of the BRDF will be covered at length.

**Colour Photometric Stereo.** The work of Vogiatzis and Hernández [50] is a member of an important class of Colour Photometric Stereo first introduced in [55]. The key idea behind these algorithms is the use of wavelength multiplexing for simultaneous acquisition of the minimum of three images for Photometric Stereo. In practice, that means that the object is simultaneously illuminated by red, green and blue light sources, which can all be read by an RGB camera sensor. Instantaneous acquisition opened the door for dynamic scene reconstruction using Photometric Stereo. Note that, unless the reflecting surface is white, with non-white illumination, albedo is also a function of the light source spectral characteristics $\mathcal{E}(\lambda)$ which introduces inconsistency in the three images acquired from different channels. With such multi-spectral acquisition, albedo thus becomes a vector due to the chromatic properties of the surface. The channel inconsistency issue initially made applicability of Colour Photometric Stereo to non-white surfaces a challenge to be solved. Brostow et al. [55] only tackle reconstruction of dynamic scenes with the assumption of either the uniformly white colour or uniform albedo across a largely Lambertian surface.
2.3. Photometric Stereo

Subsequently, the technique is extended in [56] to textured (i.e. multi-hue, spatially varying per-channel albedo) surfaces. The Lambertian reflectance limitation is addressed but not universally solved for Colour Photometric Stereo in [50] as discussed previously.

**Hybrid methods with Conventional Stereo.** One disadvantage of pure Photometric Stereo is the low frequency noise in the acquired 3D shape [50]. The artefact occurs because direct normal field integration into a continuous surface is accompanied by drift i.e. accumulation of numerical integration error. This motivated the emergence of a new class of hybrid algorithms combining Photometric and Conventional Stereo [57], [58], [59], [60] [61]. In these hybrid pipelines the depth-based Conventional Stereo acts as a shape initialiser or a merging technique for normal fields. The fine detail is added using dense Photometric Stereo. In [57], a multi-ocular photometric matching cost is proposed by combining multiview and Photometric Stereo for high-quality shape recovery under simultaneously varying viewpoint and illumination in a non-iterative procedure. In [58], the need to merge individual view-dependent normal maps is removed by applying both multiview and Photometric Stereo in a 2D parameter domain obtained using a planar mesh parametrisation technique. The combination of geometric and photometric cues in a single domain facilitates better fusion of multiview and Photometric Stereo. Wu and colleagues in [59] recover spatio-temporally coherent geometry detail to complement a coarse multiview shape in dynamic scenes by means of Photometric Stereo under general unconstrained time-varying lighting conditions. Vlasic et al. in [60] obtain mm-accurate full 3D dynamic scene reconstruction with large (human-size) working volumes at 60Hz frame rates by aligning high quality normal maps using multiview stereo constraints and thin-plane spline deformation to compensate for drift causing low-frequency deformation. Interestingly, this method starts with view-dependent Photometric Stereo reconstruction and then refines using multiview stereo rather than vice versa. Since the method places more emphasis on the photometric information using multiview only for alignment of individually integrated normal maps, the amount of texture in the test sequences can be reduced substantially (e.g. textureless garments of the reconstructed figures) without any detrimental effect on the accuracy, unlike in conventional multiview stereo methods. Since responses to eight illumination conditions are acquired sequentially in [60], optical flow is used for motion compensation of the test subject during the acquisition period. Unlike Vlasic and colleagues, Klaudiny and Hilton
in [61] also explicitly enforce temporal coherence within the reconstructed mesh sequence, specifically targeting facial expressions. In their method, a set of unaligned meshes is obtained using multiview stereo. The meshes are subsequently temporally aligned using a non-sequential tracking approach based on a minimum spanning tree built by considering similarity of pose. Lastly, high-resolution detail is added by combining per-frame normal maps obtained using Colour Photometric Stereo with the temporally consistent sequence of meshes.

The hybrid methods exploit those strengths of each method that compensate for the weaknesses of the other(s) in the union. The fusion can also be a weakness as failure of any one technique negatively impacts the whole result. Interesting research has recently been presented by Gotardo et al. in [62] on integration of Photometric Stereo, multiview Conventional Stereo and optical flow into a single energy optimisation problem for surface reconstruction. The approach called Photogeometric Scene Flow (PGSF) makes use of the premise that the result as a whole will be less susceptible to the limitations of each individual method if they are performed simultaneously supporting each other rather than sequentially which often results in categorical regularisation decisions being made too early. The presented results of extremely high-resolution temporally consistent mesh sequences of facial expressions are strong evidence that the paradigm adopted in the work is a step in the right direction for the future of hybrid methods. Such hybrid methods, as well as their older generation, do not improve robustness to complex reflectance as non-Lambertian behaviour equally poses a challenge for multiview and Photometric Stereo (as well as for optical flow).

**Full 3D dynamic scene reconstruction.** Hybrid methods have played an important role in the development of Photometric Stereo for full 3D dynamic scene reconstruction. In one prominent hybrid method [60] discussed above notably permitting full 3D dynamic scene reconstruction, Conventional Stereo matching mitigates individual biases of partial surface reconstructions induced by integration of view-dependent normals fields so that those partial surfaces can be merged into a single consistent watertight 3D model. Pure Photometric Stereo techniques for closed surface reconstruction [49,63] target static scenes only as more computational effort is required to obtain the watertight model through optimisation without the aid of geometric cues from multiview stereo. Colour Photometric
Stereo does not easily extend to full 3D dynamic scenes due to the practical difficulties to do with ensuring non-overlapping illumination intensities of sampling triplets between views [55].

**Under-constrained Photometric Stereo.** Another research direction is under-constrained Photometric Stereo. In general, for any reflectance function, Photometric Stereo requires at least three images under different illumination for each viewpoint. However, shadows and self-occlusions corrupting the minimal input image set may further reduce the count of fully usable images making the problem under-constrained. In [64] an algorithm is proposed to deal with corruption by shadows in two and three-source Photometric Stereo data. In the two-source scenario one image of the full minimal set is totally unusable while in the three-source one there is local corruption by shadows only. For shadowed pixels, normal estimation at a corrupted surface point is guided by the intersection of the remaining two constraints, which is defined by a line in 3D. The problem is formulated as a simple unconstrained least-squares minimisation jointly optimising shadowed and un-shadowed pixels using respectively a point-to-line and point-to-point model-to-data error term and appropriate regularisation. The problem is not simple and often ill-posed, with there being little decrease in the cost function for a substantial normal change. The smooth shading and smooth shape regularisation schemes can be used to combat this ill-posedness, even in the case of complete third image occlusion in the two-source Photometric Stereo scenario.

The ultimately under-constrained Photometric Stereo, when only a single image is available, is known as Shape from Shading, an independent class of reconstruction techniques reviewed in the next section.

**Assessment of Photometric Stereo techniques.** Photometric Stereo technology is highly advanced being the state-of-the-art in many industrial applications such as facial capture for film production. Unlike in Conventional Stereo, the fundamental principle of Photometric Stereo is not limited to one specific reflectance model, though it does require the model to be parametric and *a priori* known. In practice, it is difficult to both accurately describe surface reflectance as a parametric model and to obtain the parameters. In addition, an elegant mathematical integration of a non-Lambertian parametric model into the normal constraint in Equation 2.2 replacing the constant albedo can be a challenge.
As a result, the vast majority of Photometric Stereo algorithms adhere to the Lambertian reflectance model with either constant or variable albedo. It must be stressed that the need for parametric reflectance model acquisition is not simply inconvenient but may well be impossible to the desired degree of accuracy for complex real-life materials. Another issue is the fact that Photometric Stereo characterises the surface by its normals only. The only way to acquire the accompanying depths is by integrating photometric methods with Conventional Stereo. While achieving its goal under suitable conditions, that approach however also inevitably makes any such hybrid method vulnerable to the reflectance model limitations of Conventional Stereo which are more restrictive than those of Photometric Stereo.

2.4 Shape from Shading

The Principle. Shape from Shading, first introduced by Horn in the 1970s [65], can be viewed as a special case of Photometric Stereo requiring just one image for reconstruction. The method relies on the natural gradations of reflected light intensity (i.e. shading) to determine the shape of the object. The shading cue is in fact an inherent tool of the human visual system for seeing shapes. Shape from Shading shares the basic image formation equation with Photometric Stereo: \[ I(x, y) = R(p, q) \] where \( p = \frac{\partial z}{\partial x} \) and \( q = \frac{\partial z}{\partial y} \) are the normal components. In Shape from Shading however, even if the reflectance function and the illumination sources are known, the equation cannot be solved directly as it is under-determined. In Photometric Stereo, the problem is constrained by taking more images and finding the intersection of corresponding iso-brightness curves (a system of basic image formation equation is solved instead of the single equation). With the single iso-brightness curve of the Shape from Shading problem prior knowledge, typically involving some form of smoothness and shape template, is used in a cost function minimised to solve the core equation for surface normals. The data term of the formulation penalises the differences between the measured image intensity \( I(x, y) \) and the predicted \( R(p, q) \) during optimisation.

Classification. In the comprehensive taxonomy of Zhang et al. [66], Shape from Shading algorithms are grouped into four categories: minimisation-based, local, propagation-based and linearisation-based. The classification system is summarised in Table 2.3. Back in 1999, the taxonomy assesses the performance of the Shape from Shading algorithms as
2.4. Shape from Shading

table 2.3: Classification of Shape from Shading algorithms in the taxonomy of Zhang et al. \[66\].

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimisation approaches</td>
<td>prior constraints (e.g. smoothness, brightness, integrability and intensity gradient) in a global minimisation framework are introduced. piecewise-smooth surfaces can be handled too.</td>
</tr>
<tr>
<td>Local approaches</td>
<td>local shape assumptions made; said to be the most robust Shape from Shading approach.</td>
</tr>
<tr>
<td>Propagation approaches</td>
<td>propagation of shape reconstruction from a set of seed surface points (e.g. singular points) for which the normals can be easily computed.</td>
</tr>
<tr>
<td>Linearisation approaches</td>
<td>solving the image illumination equation directly by linearisation.</td>
</tr>
</tbody>
</table>

generally quite poor and identifies three main directions for development: 1. extension to non-Lambertian reflectance models; 2. extension to perspective geometric sampling and 3. combining Shape from Shading with other cues. Modern research on Shape from Shading pursues all of these directions.

**Development of modern Shape from Shading.** A more realistic perspective formulation of Shape from Shading \[67\] has come to replace the traditional orthographic one. Mathematically, the formulation exhibits depth scale invariance due to the natural logarithm of depth dependence in its image formation equation, which is a better premise than the translation invariance of the orthographic formulation. Shape from Shading is known to suffer from the so-called concave/convex ambiguity \[68\] i.e. depending on the perceived orientation of the light source relative to the surface, it will appear as either a hill or a crater. The presence of numerous solutions is a manifestation of the ill-posedness of the Shape from Shading problem. Problem re-formulations based on more realistic image formation models \[69\] have been introduced that strive towards well-posedness (i.e. uniqueness and existence of the solution) aspiring to tackle any geometry without additional information or regularity assumptions. However, methods like \[69\] tackle the geometric ambiguity only while making the assumptions of Lambertian reflectance and known positions of light sources. Like in generic Photometric Stereo, the Lambertian assumption is extremely com-
mon for its simplicity. As part of modern Shape from Shading research, the technique has been re-formulated for more realistic reflectance models: e.g. the Phong reflectance model [51] in [70]. On the other hand, the challenge of unknown arbitrary illumination given uniform but unknown surface albedo has also been tackled [71]. Further, for robustness against non-Lambertian reflectance and shape ambiguities, hybrid methods fusing Shape from Shading simultaneously with active range scanning and Conventional Stereo have been researched as well [72].

**Assessment of Shape from Shading.** The large volume of work to address the shortcomings of early Shape from Shading has led to considerable improvements in the performance of the technique. There has been a substantial improvement in the accuracy of global shape for a much wider range of objects of different geometric complexity. However, the shape obtained by Shape from Shading is often still quite coarse, lacking resolution, particularly on real datasets. Given the complexity of the under-constrained problem, research seems to target just one limitation of the method at a time e.g. geometric ambiguity, reflectance model restrictions, unknown illumination etc. Although traditionally Shape from Shading assumes Lambertian reflectance, it is possible to replace it by a more realistic reflectance model. However, just like with generic Photometric Stereo, this comes at the cost of derivational and computational complexity. In addition, the reflectance model of the reconstructed surface is assumed to be parametric and possible to estimate in practice, which may not be true. So, although the single image per frame requirement of Shape from Shading is an appealing feature for dynamic scene reconstruction, the technique does not provide enough geometry detail, freedom of reflectance and robustness to be interesting for the research question of this dissertation. However, the much improved Shape from Shading may well find its niche of suitable applications elsewhere, e.g. single camera mobile phone applications, for approximate reconstruction with extremely limited information. Such an application is particularly likely given the work on efficient algorithms [73] for Shape from Shading.

### 2.5 Structured light and related algorithms

**The Principle.** It is known that lack of surface texture is a factor that can negatively affect the performance of Conventional Stereo as correspondences cannot be easily established.
Structured light techniques address the issue by projecting light patterns onto the surface in order to artificially introduce texture. Globally these techniques can be described as depth sampling using customised illumination.

**Structured light and polarisation methods.** One of the most successful applications of structured light has been in conjunction with polarisation techniques. The roots of polarisation are in the early work on reflectance estimation [74] that introduced the Light Stage for dense reflectance sampling and techniques for diffuse/specular component separation by making use of light polarisation filters. Further work [75] made the sampling more practical and proposed a pipeline for high-resolution geometry recovery combining specular normal maps from polarised spherical gradient illumination with structured light. The ultimate application of the structured light + specular component techniques was demonstrated in the Digital Emily Project for re-animation of pore-level accurate facial models. The low-resolution geometry is acquired by projection of colour fringe patterns for stereo correspondences. The high-resolution geometry is obtained from polarised spherical gradient illumination [75] by subtracting cross-polarised from parallel-polarised images for diffuse/specular separation and computing the specular normals from four gradient illumination patterns. Fusion of high and low resolution models gives a globally accurate high resolution geometry. Ghosh et al. [76] generalise the pipeline to multiview and show how to obtain low-resolution geometry from diffuse/specular information directly without resorting to structured light. Polarisation techniques are limited to non-metallic surfaces as metals do not polarise light on reflection.

**Structured light and depth map fusion.** One immensely successful structured light implementation using commodity hardware is KinectFusion [77], [78]. An interesting precursor of the work is another structured light approach [79] for real-time interactive 3D model acquisition. Rusinkiewicz et al. [79] project spatio-temporally-varying patterns of stripes using visible light. As explained in [79], pattern beaming for depth resolution works by finding the intersection between the camera ray to the point and the light sheet of the stripe illuminating the point. Methods tracking stripe boundaries over time have been developed to allow for (rigid) object movement within the temporal window of one pattern sequence. Range images obtained by the discussed structured light techniques are aligned using the Iterative Closest Point algorithm (ICP) [80]. Rusinkiewicz et al. developed an
efficient ICP variant with the point-to-plane error metric: scans are aligned by minimising the distance between a point on one scan and the plane through the projected point on the other scan perpendicular to the normal at the point. The user is responsible for view planning to fill in any missing information.

The structured-light-based transceiver used in Kinect technology for instantaneous depth map acquisition is infra-red, which makes it unaffected by variable lighting unlike RGB (visible light) sensors. The algorithm for geometry modelling employed by KinectFusion is closely related to work on Structure from Motion [81], [82]. The key strategy of such model building is computation of accurate geometry from many instantaneously acquired depth maps that are individually poor in quality. The model is built by gradual information accumulation and continuous model refinement. Unlike all the other methods presented in this literature survey, Structure from Motion approaches do not assume that the cameras are a priori geometrically calibrated.

Structure from Motion originates from the early Simultaneous Localisation and Mapping (SLAM) in robotics [83] - an iterative technique to simultaneously optimise the map of scanned environment and the robot’s position in it. Mapping and localisation are prerequisites of each other and, if neither is known, their inference must be joint [83]. Pollefeys [84] introduced and developed Structure from Motion in computer vision. Parallel Tracking and Mapping (PTAM) [81] and Dense Tracking and Mapping (DTAM) [82] are the recent Structure from Motion variants. Tracking and mapping, the two complementary tasks intimately linked in classical SLAM, are decoupled in PTAM meaning that data associations are no longer shared resulting in the freedom of the tracking method choice and parallel scheduling of tracking and mapping tasks for efficiency. Re-mapping need not any longer happen every frame. Bundle adjustment (i.e. the re-projection error minimisation) re-maps integrating new information at key frames and simultaneously optimising the camera position. Unlike PTAM with its sparse feature point matching, DTAM aligns views based on the totality of the available data and is more robust against occlusions, motion blur and other issues causing track loss in point-based systems.

KinectFusion is inherently not suitable for dynamic scene reconstruction since it shares this accumulative nature of Structure from Motion with the basic idea of fusing together the raw, highly incomplete depth maps generated by the Kinect sensor into a complete model.
Some algorithms have been proposed to bypass this limitation. In [78], the algorithm relies on an a priori computed static background reconstruction to segment the dynamic scene component and reconstruct it in a separate volume and accumulative procedure. Weighted accumulation giving the preference to new samples can to some extent sustain reconstruction of a continuously moving object. Memory can also be a limitation for reconstruction of larger scenes using KinectFusion due to its data accumulative nature resulting in the need for a more effective model representation. To prevent memory from affecting applicability of KinectFusion, coarse-to-fine octree-based implementations [85] as well as the moving-volume KinectFusion [86] have been proposed. The latter permits a gradual scene exploration without committing it to dynamic memory in its entirety. In [87], the volumetric representation is replaced by a memory-efficient point-based one allowing spatially extended reconstruction. As the point-based representation is native to Kinect sensor, the computational overhead in conversion is reduced in the work. Keller et al. also claim their representation allows to more easily insert or extract objects to or from the set of points used for camera pose estimation hence facilitating dynamic scene segmentation. Although an excellent progress has been made in extending KinectFusion to dynamic scenes, none of these methods can claim to have succeeded fundamentally which is actually impossible without dispensing with the assumption of static geometry.

There is much ongoing research on the topic of optimal fusion of partial incomplete 3D scans (such as the depth maps generated by the Kinect) both spatially and temporally while allowing for scene deformation [88], [89], [90]. This new generation of approaches targets 4D model acquisition explicitly whereas, in marked contrast to that, KinectFusion is inherently limited to 3D model acquisition assuming temporal stability of geometry. The spatio-temporal fusion of partial 3D observations into a 4D model effectively amounts to dynamic scene reconstruction. Let us elaborate on some interesting recent examples of such approaches. Guo et al. target dynamic reconstruction from a single-view depth input concentrating on the tracking aspect. Their observation of the motion energy being confined to distinct epicentres on the mesh led them to propose a motion regulariser that inherently seeks such pivotal points of motion thus reducing the solution space of the tracking problem. There is a skeletal-based aspect to this method but it does not require a template explicitly. The authors claim that the indirect skeletal assumption allows to generalise to mixed scenes
involving elements that cannot be modelled by skeletons. Yet, all their datasets in some way involve the human body that perfectly conforms to the skeletal assumption of their approach with joints being the epicentres of motion energy. One may wonder how the technique would cope with scenes containing less articulated motions, e.g. fabric deformation, without any human presence at all. On the other hand, Xu et al. [89] propose a template-less approach that learns the deformation model from the partial temporal observations. The deformation model is continuously updated with every new incoming observation together with the shape and allows to predict the hidden parts in any given frame. Truly a culmination in this line of research was the recent development of DynamicFusion [88] said to fundamentally address the challenge of KinectFusion generalisation to dynamic scenes. DynamicFusion is also an example of spatio-temporal fusion for 4D model acquisition and has been shown to reconstruct non-rigidly deforming scenes in real-time using a single depth camera. The basic idea of the system consists in having a canonical model space representing the starting point of the deformation and a warp field describing the motion relative to the space. Every new frame the model-to-frame warp field parameters are estimated and the frame with any information it has got to contribute to the canonical model is fused into the sequence expressed in the canonical space. The warp field is extended to accommodate the newly added to the sequence geometry evolution. Ultimately, like with all such 4D fusion approaches, the knowledge of the warp field throughout the sequence allows to compensate for noisy or incomplete data at any given time instant using the information from all the other frames: the initial canonical model can be observed to grow, becoming more complete, due to the information accumulated throughout the sequence which translates into better completeness at any given subsequent position in the sequence as the canonical model can be warped to it. Like KinectFusion, DynamicFusion employs the principle of data accumulation over time but where KinectFusion assumes static geometry used to track a moving depth camera, DynamicFusion, to build the canonical model, explicitly seeks frame-to-frame geometric correspondences in the temporally non-rigidly evolving scene by computing motion flow. The warp-field in [88] serves the same purpose as the subspace representation of object deformation in [89]. DynamicFusion is an impressive dynamic scene reconstruction method based on Kinect depth maps with a highly desirable real-time aspect. However, like KinectFusion, Dynam-
2.5. **Structured light and related algorithms**

icFusion still remains an accumulative system which brings with it the following inherent limitations. Firstly, there is a limit on the speed of scene evolution that can be handled as the algorithm needs time to accumulate the model, requiring user interaction for view planning to fill in any holes by either moving the camera to acquire more possible unseen data (KinectFusion) or by introducing scene deformation to be tracked for topology data (DynamicFusion). Any topological changes to the scene take time to fully manifest themselves in the sequence reconstructed by DynamicFusion. Similar to the reconstructed volume size restrictions of KinectFusion, in DynamicFusion there are likely to be limitations on the length of the dynamic sequence and the extent of deformation with respect to the canonical model that can be processed. In addition, while the key strength of KinectFusion/DynamicFusion is the speed of acquisition with the immediate feedback to the user, the resolution of final models is often found lacking and the reconstructions primarily represent the global shape only. There are also some characteristic reconstruction artefacts such as the inability to reconstruct thin structures and the rounded corners appearance.

**Laser-based.** For completeness one should mention the class of techniques which uses laser light, yet another form of sampling illumination next to white, polarised and infra-red light. Laser scanners can be loosely speaking classed as structured light as the laser beam dots are purposely projected onto the surface. Levoy *et al.* [91] developed impressive technology using both triangulation and time-of-flight laser scanners to digitise appearance of antique statues down to the chisel mark. Triangulation scanners [92] compute depth based on the projected laser feature (dot or stripe) and its location in the camera image domain knowing the distance between the camera and the laser and the orientation of the two relative to the laser feature projected in the 3D world. There is a noticeable similarity of the technique to the Kinect sensor that projects an infra-red pattern of dots and assesses the response received on a monochrome CMOS sensor placed at an offset from the infra-red light source.

**Assessment of structured light techniques.** Due to the great versatility of structured light techniques, one cannot fully generalise with respect to the strengths and weaknesses of the class as a whole. One unifying structured-light characteristic is that, by introducing easily detectable features artificially, the methods reduce the dependence on the presence of native scene texture and sensitivity to the choice of reflectance model - two factors that degrade the performance of feature-based and intensity-based Conventional Stereo. Complexities of
reflectance behaviour will affect the methods only in the extreme case when they prohibit visibility of the projected feature. However, since the inherent mechanism of these structured light techniques is sparse-feature-based, dense high-resolution model reconstruction in this case requires either the assistance of another technology (e.g. polarisation) or large amounts of input information accumulated gradually (e.g. KinectFusion/DynamicFusion, laser scanners). Let us assess the discussed structured light technologies individually.

The hybrid methods combining structured light with polarisation are known to produce globally accurate and extremely high resolution results. Their performance in graphics applications, such as the Digital Emily character re-animation, is striking in its realism. Unfortunately, the light polarisation technique, responsible for the remarkable level of detail, is not universally applicable because some materials (e.g. metals) do not polarise light on reflection. Without fusion with polarisation, the structured light algorithm projecting a fringe pattern will produce a globally accurate but comparative low resolution model. In addition, one should mention the practical complexity of the method’s acquisition procedure requiring a professional Light Stage with over 100 LED lights.

Laser-based structured light techniques face similar objections with their highly specialised motorised equipment tailored to the object’s dimensions as required for resolution of fine surface structure. While the triangulation-based laser scanners are accurate, they are limited in range to a couple of meters (typically the accuracy is about 50 µm for the arm mounted and rotation stage technologies at the operating range of 0.1 – 1 m [93]). In certain situations it may be challenging to have the equipment sufficiently close to the object. The much longer range (up to hundreds of meters) time-of-flight scanners suffer from imprecise light round-trip measurements affecting the accuracy of depth estimation (3 – 6 mm [93]). Naturally, laser scanning is not suitable for reconstruction of dynamic scenes due to the long latencies of the acquisition process. Any dynamic behaviour during acquisition will distort all the accumulated data.

The Kinect-based technology is on the opposite end of the spectrum with the acquisition procedure requiring only an inexpensive Kinect sensor. Despite the wide-spread popularity, KinectFusion and the emerging DynamicFusion have got limitations. First of all, the resolution of KinectFusion/DynamicFusion models is low compared to other structured light techniques or Photometric Stereo. Secondly, they are accumulative meaning that they have
inherent difficulties with dynamic topological changes in the scene. KinectFusion cannot deal with dynamic scenes at all. DynamicFusion addresses that limitation of KinectFusion but only given a sufficiently gradual scene evolution and without instantaneous reaction to topological change (i.e. objects appearing or disappearing). DynamicFusion cannot at this time compete with, for instance, Colour Photometric Stereo in terms of accuracy, resolution or the allowed speed of deformation in dynamic scene reconstruction, although the technique is characterised by the unique real-time processing highly desirable in many applications.

2.6 3D reconstruction of strongly specular surfaces

There has also been research specifically targeting reconstruction of surfaces characterised by particularly difficult optical phenomena such as the purely specular, mirror-like reflectance, discussed in this section.

Three classes of techniques specifically designed for the purpose of reconstruction with purely specular reflectance are Shape from Specular Flow [94], Shape from Distortion [95] and Shape from Specularities [96], [97]. Shape from Distortion methods derive information from the way a highly specular surface deforms a pattern projected onto it. This approach typically requires a substantial number of images. Tarini et al. in [95] use a simple CRT colour monitor to project patterns onto mirroring surfaces captured by a single high quality camera. Shape from Distortion in this case operates through an optimisation initialised by random depth estimates run on combined principles of Shape from Shading (i.e. normals from reflectance patterns) and triangulation-based techniques as for mirroring surfaces the normals bisect one of the angles in the light source-point-camera triangle. The distortion of the projected patterns in the acquired image is determined by the normal belonging to the point of intersection between the incoming ray and the surface. Depths and normals are optimised jointly by the method. Shape from Specular Flow [94] derives geometry from the propagation of specular highlights across the surface given known camera motion and distant unknown illumination. These techniques make the smooth surface assumption and, as all flow-based systems, are prone to breaking on textureless surfaces or due to occlusions. Instead of projecting a pattern like in Shape from Distortion, Shape from Specularity approaches use native highlights from an ordinary light source to reconstruct. For example,
knowing the position of the light source, Chen et al. [96] translate the resultant specularity field into the corresponding normal field to compute fine surface structure. If the specularities are localised, diffuse features need to be incorporated into the system [97]. Only local surface patches are reconstructed in both [96] and [97].

Building on the earlier work, a new type of probabilistic geometry estimation methods appeared [31], [98]. In [98] mirror-like, purely specular surfaces are targeted specifically as opposed to the more mixed, diffuse-specular surfaces in [31]. The key concept of the approach in [98] is that, given purely specular reflectance, the colour of any surface point is not an inherent property but in fact a reflection of the surrounding environment. The incident light field is thus modelled as an environment map i.e. a 3D function mapping incoming ray directions to colours of the environment’s components. Due to the purely specular reflection, the colours observed on the reconstructed mirror-like surface depend only on the surface normal and the position of the camera. The method is probabilistic because a distribution of possible normals per vertex is built. The optimisation is initialised with the visual hull, which is already a fair representation as about 30 images from different viewpoints are typically used. In an optimisation procedure converging in under 10 iterations, the normals are iteratively inferred per vertex with the local smoothness assumption and the mesh is adjusted accordingly. Another recent work [99] proposes an interactive method for reconstruction of strongly specular and transparent objects based on the reflectance-independent visual hull. Zuo et al. in [99] tackle the problem of concavity reconstruction, fundamental for Shape from Silhouette, by relating every concavity to detected convex points of the internal contours. The basic idea is that every surface “valley” should be flanked by “hills” i.e. convex surfaces. The global optimisation procedure is developed for internal contour tracking from a set of keyframes where the user manually marks the internal contours. Although typical data terms are used in the formulation of the energy minimisation problem, additional regularisation terms are added to make the formulation more robust for the targeted specular or transparent surfaces for which the assumptions of the data terms may not always hold. Impressive results have been achieved with this method, however it has got limitations. Firstly, it is highly dependent on user interaction at several stages of the pipeline. Secondly, contour tracking on the targeted challenging surfaces is often not robust enough when the smoothness terms cannot regularise the unpredictable observations.
Essentially, weak Lambertian assumptions are made at the tracking stage of the pipeline meaning that the proposed method is not fundamentally suitable for the specular model. Finally, the system is reported to under-fit the concavities, which is not surprising since they are based on a set of estimated convex points of the internal contour albeit refined using the smart Locally Convex Carving procedure.

In [100] a comprehensive overview is given of the techniques specifically targeting reconstruction with complex optical behaviour which is out of scope for most mainstream reconstruction techniques. Apart from the discussed mirror-like reflectance, these include reconstruction in the presence of transparency with refractive behaviour or severe translucency/sub-surface scattering. As the concern for sub-surface optical phenomena is out of this dissertation’s scope (except as a source of error), this section concentrates on reviewing and assessing strongly specular reconstruction only in this class of methods targeting reconstruction with outlier optical behaviour.

**Assessment of the techniques targeting strongly specular surfaces.** These methods with out-of-the-box thinking designed to handle extremely challenging reflectance behaviour are impressive, in particular the recent probabilistic geometry estimation approaches with environment maps. For example, in their recent work Godard et al. [98] accurately reconstruct non-trivial full 3D shapes with mirror-like surfaces which is impossible by any established method. However, their technique is limited to mirror-like surfaces only as purely specular reflection is assumed. An earlier related technique in [31] allowing a more realistic model with both diffuse and specular components does not achieve the same level of accuracy on real datasets. Similarly, the earlier approaches also rely on specific (extreme) reflectance behaviour. Shape from Specular Flow assumes the object is substantially and uniformly specular for reliable tracking of specularities. Shape from Distortion assumes the distorted pattern will actually show on the acquired image which suggests mirror-like behaviour. Even if the pattern can be detected on other specular surfaces, the performance will be limited as the key equations in the method are based on the assumption of the surface normal bisecting the angle formed by the incident and reflected rays, only absolutely true for mirror-like surfaces. Shape from Specularity, like Shape from Specular Flow, requires a consistent distribution of specularities across the surface.

So, as such, all these strongly specular behaviour techniques do not solve the problem of
reconstruction with an arbitrary reflectance model but rather propose methods specifically tailored to mirror-like reflectance. Some of the techniques would not work on standard Lambertian surfaces and even those with a realistically heterogeneous diffuse-specular reflectance model.

### 2.7 Helmholtz Stereopsis

**The Principle.** Helmholtz Stereopsis [101] is a comparatively recent reconstruction technique that addresses the fundamental reflectance modelling problem in geometric reconstruction by enforcing consistency of the reflectance-model-independent Helmholtz reciprocity. The technique works with arbitrary reflectance models resting on neither specular nor Lambertian behaviour of the surface unlike the methods presented in the previous section and Conventional Stereo respectively. Helmholtz Stereopsis is classified in [102] as a photo-geometric technique as it utilises both a special acquisition geometry like the geometric methods (Section 2.1) and photometric measurements of reflected illumination like the intensity-based methods. Helmholtz reciprocity [103] states that a light ray and its reverse will undergo identical optical processes. Let $v_1$ be the unit vector directed from the surface point to the camera and $v_2$ the corresponding vector from the surface point to the light source as shown in the left image of Figure 2.1. The implication of Helmholtz reciprocity, first observed by Zickler et al. [101] in the context of geometric reconstruction, is that interchanging the light source and the camera in the set-up, as in the right image of Figure 2.1, has no effect on the point’s reflective behaviour. The BRDF $f_r$ is reciprocal:

$$f_r(v_2, v_1) = f_r(v_1, v_2)$$ (2.4)

The image formation equations for reciprocal images $I_1$ and $I_2$ respectively:

$$i_1 = \rho_{iso} f_r(v_2, v_1) \frac{n \cdot v_2}{\|c_2 - x\|^2}, \quad i_2 = \rho_{iso} f_r(v_1, v_2) \frac{n \cdot v_1}{\|c_1 - x\|^2}$$ (2.5)

express image intensities $i_1$ and $i_2$ of surface point $x$ as a function of BRDF, surface normal $n$, the two reciprocal unit vectors, the camera/light source positions $c_1/c_2$ and the spatially constant radiance distribution $\rho_{iso}$ of an isotropic point light source assumed. In [101],
the same isotropic radiance distribution is assumed to be used for acquisition of both \( I_1 \) and \( I_2 \). By substitution of re-arranged intensity expressions from Equation 2.5 into the BRDF equality in Equation 2.4, one obtains the following constraint \( w \) notably \textit{without} any dependency on the BRDF:

\[
\left( \frac{v_1}{\|c_1 - x\|^2} - \frac{v_2}{\|c_2 - x\|^2} \right) \cdot n = w \cdot n = 0. \tag{2.6}
\]

Note also that the constant isotropic radiance distribution \( \rho_{iso} \) cancels out being the same in both image formation equations. With one \( w \) per reciprocal pair, three or more reciprocal pairs result in constraint matrix \( W \) suitable for Singular Value Decomposition (SVD):

\[
\text{SVD}(W) = U \Sigma V^\top \tag{2.7}
\]

where \( U, V \) are orthogonal and \( \Sigma \) is a diagonal matrix. The last column of \( V \) gives the normal at the sampled point. The last diagonal value of \( \Sigma \), the SVD residual \( \sigma_3 \), tends to zero when there is a mutual consistency of constraints. For outlier elimination, Zickler \textit{et al.} \[101\] formulate the consistency measure as the quotient \( \frac{\sigma_2^2}{\sigma_3^2} \), which tends to infinity for true surface points.

Helmholtz Stereopsis is formulated as a \textit{maximum likelihood} problem. A reconstruction volume containing a set of sought surface points is swept, computing a normal for each depth hypothesis. For each surface point sought, the depth hypothesis with the highest consistency coefficient \( \frac{\sigma_2^2}{\sigma_3^2} \) is chosen independently of all the other points. As a consequence unlike all the previously described methods each surface point in Helmholtz Stereopsis is characterised by both a depth and a normal.
Further developments. Helmholtz Stereopsis in its original formulation imposes the strict requirement of the minimum of three reciprocal pairs and careful offline calibration. Calibration. The requirements had led to discussions on impracticality of the technique’s acquisition. In response Zickler in [104] devises an auto-calibration algorithm using the inherently identifiable specular highlights and intensity patches. Both geometric and basic radiometric calibration can be thus obtained. By virtue of reciprocity, specularities will be located at corresponding image points in any two reciprocal images. Unlike Conventional Stereo, intensity matching in Helmholtz Stereopsis is not conditional on the validity of the Lambertian assumption making intensity patches stable calibration markers as well: a locally planar patch in the right image of the pair will be identical to the corresponding patch in the left image up to a scale factor. This ability to predict an image from its reciprocal has also been employed in registration for full 3D Helmholtz Stereopsis [105].

In [104], the definition of radiometric calibration is limited to measuring relative intensities of isotropic light sources. Provided the assumptions of equal camera responses and no spatial illumination intensity variation hold, such limited calibration is sufficient. A more general radiometric set-up calibration for Helmholtz Stereopsis, which does not rely on these assumptions, was proposed by Jankó et al. [106]. Using a sequence of localised calibration planes Jankó et al. calibrate for a spatially varying joint parameter describing sensitivity and radiance of a collocated camera and light source pair.

Full 3D Helmholtz Stereopsis. The work on full 3D Helmholtz Stereopsis has in fact been scarce. Delaunoy et al. [107] and the more recent Weinmann et al. [108] are two notable examples. Both are variational approaches requiring computationally intensive time-consuming optimisation over the entire surface. The method in [108] is a hybrid approach involving fusion with structured light at acquisition. The increased complexity, as the result of that fusion in [108], for certain applications would be justified by the impressive degree of demonstrated reconstruction detail. As any hybrid approach, the method in [108] seeks to exploit those strengths of each technique that would compensate for the weaknesses of the other: essentially, in this method the global shape must be provided by structured light as depth estimates by conventional Helmholtz Stereopsis are typically noisy, whereas the dense normal field, accurate regardless of the reflectance model, can be estimated using the Helmholtz reciprocity principle. The method of Delaunoy et al. in [107] on the contrary is
Based on pure Helmholtz Stereopsis. Aside from pioneering full 3D reconstruction using the technique, it seems to provide some resolution improvement compared to the conventional volumetric Helmholtz Stereopsis first proposed by Zickler et al. (no comparative evaluation is presented in the paper however). However, its variational formulation is computationally intensive and even the 2.5D reconstruction results presented seem to require $8-18$ reciprocal pairs. So, despite the improved accuracy, the branch of variational Helmholtz Stereopsis is not suitable for dynamic scene reconstruction. The method of Zickler et al. remains the state-of-the-art volumetric Helmholtz Stereopsis.

**Number of reciprocal pairs.** To address the requirement of at least three reciprocal pairs, which may be inconvenient in some cases, Zickler et al. [102] propose binocular Helmholtz Stereopsis. The method is a differential approach where a single constraint is formulated as a Partial Differential Equation (PDE) of depth over surface coordinates. The PDE requires initialisation and results in a family of solutions, disambiguated through a multi-pass optimisation. Thus the significantly increased computational complexity is traded for simplification and speed-up of acquisition.

**Formulation.** Guillemaut et al. [109] provide an alternative formulation of the Helmholtz Stereopsis constraint. Their formulation is called *radiometric* because it expresses the constraint in terms of intensity modification. Not only is such a formulation more intuitively transparent than Zickler’s original *algebraic* formulation but, being a sum of squared errors, it also lends itself to the optimal solution by *Maximum Likelihood (ML)* minimisation. The optimality guarantee is only valid under the assumptions of 1. intensity errors dominating over any calibration errors and 2. intensity measurement errors being statistically independent and Gaussian. Experimental results of Guillemaut et al. have shown that, on synthetic data with the noise model exactly confirming to the made assumption, radiometric formulation solved by ML minimisation consistently outperforms algebraic SVD-based approaches. On the real data, the two are very close in performance, with the algebraic measure approaching the ML estimate.

Guillemaut et al. in [109] raise another important issue rarely addressed in research on Helmholtz Stereopsis, namely that of image saturation. When saturation of the camera sensor occurs, one can no longer reliably compute the constraint, whatever its formulation. Even though not solving the problem fundamentally, the authors propose a measure for its
mitigation by assuming that saturation occurs in both images of the pair simultaneously. Under this assumption, the unknown intensity value in such a saturated reciprocal pair can be replaced by an arbitrary pixel saturation intensity value. At such saturation points, the normal bisects the incident and emergent rays, allowing to further simplify both algebraic and radiometric constraints. Although effective in some cases, the approach will fail when the underlying assumptions are invalid. Reconstruction by Helmholtz Stereopsis in the presence of image saturations is therefore still an open topic for research.

Furthermore, Guillemaut et al. in [110] re-consider intensity sampling in Helmholtz Stereopsis. On rough or textured surfaces, intensity sampling in a single reciprocal pair may be perturbed due to inter-reflections and/or geometric calibration inaccuracies: a single point may be projected into non-corresponding intensity fields in the two reciprocal images. To mitigate the effect, instead of the intensities at the point of projection alone, averages over local neighbourhoods of the projection in the image domain are used. Such an averaging approach acts to increase the degree of correspondence in the sampled intensity fields in the pair of reciprocal images. For rough surfaces this facilitates the recovery of the macro-structure, despite the inter-reflections caused by its micro-structure, whereas, in the presence of surface texture, the approach smooths out any artefacts at the intensity region boundaries.

Assessment of Helmholtz Stereopsis. Helmholtz Stereopsis is a relatively recent but already very promising technique in the world of 3D geometric reconstruction. The key advantage of Helmholtz Stereopsis over earlier methods is its complete independence of the reflectance model by virtue of its acquisition set-up. Another unique feature of Helmholtz Stereopsis is its ability to characterise surface points by both depth and normal estimates. The property is highly desirable as one-sided characterisation by depths or normals alone has been known to result in artefacts (e.g. the lacking resolution of depth-based Conventional Stereo or the drift-prone global shape of Photometric Stereo). Conventionally, however, Helmholtz Stereopsis does not use the depths, processing the reconstruction output purely as a normal field like in Photometric Stereo.

Although some work has already been done on improving Helmholtz Stereopsis, there are still a few limitations prohibiting its widespread utilisation for 3D reconstruction with complex reflectance. Firstly of all, the original acquisition paradigm requiring sequential
inter-changing of the camera and the light source is not very practical, especially seeing that for accurate reconstruction up to 18 reciprocal pairs (i.e. 36 images) were needed in the seminal paper [101]. The sequential acquisition prohibits reconstruction of dynamic scenes, even if the under-constrained Binocular Helmholtz Stereopsis is used at the expense of accuracy. Alternative variational formulations, in addition to computational complexity, suffer from the same limitation of high reciprocal pair requirement as the original volumetric Helmholtz Stereopsis. The high reciprocal pair count suggests sub-optimality of formulation, which manifests itself at a lower reciprocal pair count by extreme noisiness of the estimated depth map. The noise is likely to be the reason the estimated depths are not utilised in Helmholtz Stereopsis. The utilisation of depth estimates is important to support accuracy of global shape reconstruction.

Despite the shortcomings of the state-of-the-art Helmholtz Stereopsis, the technique is extremely promising for the research question of this dissertation due to its property of reflectance model independence. Since dynamic scenes are targeted, Helmholtz Stereopsis will need to be generalised beyond static scene reconstruction. The original volumetric Helmholtz Stereopsis by Zickler \textit{et al.} is more promising for such an extension than the variational variants and is considered the state-of-the-art in this work. As the current reconstruction accuracy of Helmholtz Stereopsis at lower reciprocal pair counts is unsatisfactory, its formulation will need to be re-considered. One idea developed in this thesis is the inadequacy of the traditional formulations being caused by the failure to integrate depth and normal information at different stages of the pipeline, essentially wasting one of the unique advantages of Helmholtz Stereopsis. Another perceived shortcoming in the formulation to be addressed is the maximum likelihood nature of the state-of-the-art volumetric Helmholtz Stereopsis, which in practice means that the reconstructed surface is treated as a collection of independent points rather than a continuous entity.

\section*{2.8 Summary and discussion}

Table 2.4 compares and contrasts the performance of the 3D reconstruction algorithms discussed in this chapter. The methods from Section 2.6 for mirror-like surface reconstruction are purposely left out of consideration as they will not work on surfaces with any other reflectance behaviour. In terms of geometric quality several methods perform well given
tailored surface properties. In fact, only Shape from Silhouette and Shape from Shading can be discarded straight away on the grounds of coarseness of geometry and a highly inconsistent quality of geometry respectively. In the pool of remaining algorithms with moderate to extremely high geometric accuracy, many are limited in their scope to surfaces with some specific reflectance properties. Conventional Stereo assumes Lambertian reflectance whereas Photometric Stereo needs an \textit{a priori} known reflectance model that can be parametrised. Structured light with polarisation cannot handle metallic surfaces. Further, although KinectFusion is robust to Lambertian assumption violations due to the accumulative nature of the model building algorithm, it is still susceptible to errors when the projected infra-red pattern cannot be detected due to specularities. The resolution of KinectFusion is also quite low and, although it has recently been extended to dynamic scenes in its new re-incarnation as DynamicFusion, the technology is still inherently not well-suited to efficient reconstruction of arbitrary speed dynamic scenes, due to its spatio-temporally accumulative nature.

From the techniques with acceptable geometric accuracy and resolution, only laser-based structured light and Helmholtz Stereopsis do not impose reflectance model limitations. Laser-based structured light cannot be used for dynamic scene reconstruction and is also impractical because of its dependence on highly specialised scene-scale tailored equipment. Compared to that, Helmholtz Stereopsis uses standard acquisition equipment of cameras and light sources. The technique has not yet been generalised to dynamic scenes but there is no limitation in its paradigm itself fundamentally preventing that. With the unique properties of reflectance model independence and dual characterisation of the surface by both depths and normals as well as the potential for generalisation to dynamic scene reconstruction, the till now overlooked technique of Helmholtz Stereopsis can be developed to become competitive in the field of accurate high-resolution geometric reconstruction of dynamic scenes with complex reflectance properties.
<table>
<thead>
<tr>
<th>Method</th>
<th>Geometry quality</th>
<th>Reflectance (BRDF) limitations</th>
<th>Other limitations</th>
<th>Suitable for dynamic scenes?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shape from Silhouette</td>
<td>depends on the number of images available; mostly coarse; concavities cannot be reconstructed</td>
<td>none</td>
<td>strong dependence on the number of views;</td>
<td>yes</td>
</tr>
<tr>
<td>Conventional Stereo</td>
<td>accurate global shape and moderate resolution</td>
<td>Lambertian reflectance model</td>
<td>1. surface texture needed; 2. surface characterisation by depths only</td>
<td>yes</td>
</tr>
<tr>
<td>Photometric Stereo</td>
<td>drift-prone global shape and high resolution</td>
<td>any a priori known parametric reflectance model (typically Lambertian for simplicity)</td>
<td>1. surface characterisation by normals only; 2. complexity of non-Lambertian formulation; 3. difficulty of a priori reflectance model estimation and parametrisation</td>
<td>yes</td>
</tr>
<tr>
<td>Shape from Shading</td>
<td>coarse and/or low resolution</td>
<td>the same as Photometric Stereo</td>
<td>the same as Photometric Stereo + 4. inherently under-constrained problem (priors needed)</td>
<td>yes (potentially)</td>
</tr>
<tr>
<td>Structured light and polarisation</td>
<td>extremely accurate global shape and remarkable resolution</td>
<td>non-metallic surfaces</td>
<td>1. needs the Light Stage, specialist carefully calibrated expensive acquisition equipment</td>
<td>yes</td>
</tr>
<tr>
<td>Structured light and depth map fusion (KinectFusion/DynamicFusion)</td>
<td>accurate global shape and moderate resolution</td>
<td>Lambertian preferred but with a high degree of tolerance to violations</td>
<td>1. cannot be used outdoors; 2. online user interaction required in view planning</td>
<td>yes (with limitations)</td>
</tr>
<tr>
<td>Laser-based structured light</td>
<td>extremely accurate global shape and remarkable resolution</td>
<td>none</td>
<td>1. highly specialist bulky and expensive equipment; 2. range limitations of the triangulation-type laser-based techniques; 3. time consuming acquisition</td>
<td>no</td>
</tr>
<tr>
<td>Helmholtz Stereopsis</td>
<td>dependent on the number of reciprocal pairs; accurate global shape and moderate resolution; surface characterisation by depths and normals</td>
<td>none</td>
<td>1. about 10 – 20 reciprocal pairs typically required for geometric accuracy; 2. sequential acquisition</td>
<td>not yet</td>
</tr>
</tbody>
</table>

Table 2.4: A comparative overview of the state-of-the-art reconstruction techniques in terms of geometric quality, reflectance model (BRDF) limitations and applicability to dynamic scenes.
Chapter 3

Coarse-to-fine Bayesian Helmholtz Stereopsis
Active research in the field of 3D reconstruction has produced a variety of methodologies yielding sub-millimetre accurate geometries when capture conditions and surface properties are tailored for reconstruction. Conventional \( [10, 12, 13, 24, 28, 35, 36] \) and Photometric \( [45,49,55,56,60,63] \) Stereo are two examples of mature techniques that have demonstrated exceptional results within the widest application scope. Certain underlying assumptions however prevent these techniques from being universally applicable. Conventional Stereo relies on uniformly Lambertian (purely diffuse) surface reflectance for consistency of feature point appearance needed to match them between views. The reliance is restrictive as purely Lambertian photometric behaviour is not common for real objects. Photometric Stereo has no restrictions on the reflectance behaviour type if the model linking the surface normal to directional illumination response is known \textit{a priori} and is parametric. Complexity and spatial variation in real reflectance models often make them impossible to estimate and parametrise resulting in the Lambertian assumption commonly being made by photometric techniques as well. More recent structured light techniques similarly still impose limitations on the surface type (e.g. non-metallic material) or the content of the scene (e.g. the difficulty KinectFusion has with dynamic scene reconstruction).

The classical Shape-from-Silhouette \( [1] \), a geometric technique unfortunately limited due to the fundamental inability to reconstruct concavities, had held the monopoly on reflectance model independence. HS \( [101] \) was the first intensity-based method independent of the reflectance model by virtue of its clever normal constraint formulation with a tailored acquisition set-up. The technique builds on Helmholtz reciprocity which allows to eliminate reflectance from the equation. The normal constraint is computed based on sets of reciprocal pairs of images acquired with the mutually inter-changed camera and light source. The validated performance of standard HS however relies on many reciprocal pairs (e.g. 18 pairs \( 36 \) images in the seminal paper \( [101] \)) and even then is known to produce noisy depth maps. Reconstruction is pulled together by direct integration of the estimated normal filed completely disregarding the corresponding depth estimates. In short, being revolutionary in the key concept, the method is in need of 1. a better formulation allowing it to reach its full potential playing on the strengths of the technique (e.g. surface characterisation by both depths and normals) and 2. a simplification of the reciprocal pair acquisition procedure in terms of latency per reconstructed (static) shot. The simplification can be accomplished
by reducing the number of reciprocal pairs required per frame for reconstruction of desired quality and/or by decreasing the capture time per reciprocal pair ultimately striving for simultaneous acquisition.

In this chapter, to address the first shortcoming of standard HS, a novel Bayesian formulation of HS is developed to replace the original sub-optimal maximum likelihood (ML) formulation. The Bayesian formulation by definition entails neighbourhood support in solving the reconstruction problem. Further to introducing neighbourhood support, the most unrestricted prior, enforcing surface integrability only, is designed to encapsulate it. This integrability prior is tailored to the strengths of HS, specifically to the unique ability of the technique to generate per vertex both depth and normal estimates. Further adjustments to the pipeline include elimination of the traditionally used surface integration method to promote consistency between the generated continuous surface and the point cloud outputed by the reconstruction algorithm. The consistency without any optimising post-processing is desirable in this context due to the exceptional quality and resolution of the generated point clouds facilitated by the tailored optimal integrability prior and the Coarse-to-fine (CtF) framework the pipeline is embedded into. In summary, the contribution of the chapter is the CtF Bayesian HS - the novel Bayesian formulation of HS together with all its tailored features resulting in a fundamentally improved quality of reconstruction with a much reduced reciprocal pair count as input.

The chapter is organised as follows. In Section 3.1 the problem of 3D reconstruction is formalised with the notation and terminology best applicable to HS and Bayesian problems solved by MRF optimisation. The theory of state-of-the-art HS is introduced in Section 3.2.1 followed by the methodology of the proposed Bayesian HS in Section 3.2.2 including a comparative discussion of the different priors applicable in the formulation. Section 3.3 seeks to justify the choice of prior for Bayesian HS relating it to surface integrability as the most optimal least restrictive regularisation assumption. The implementation is covered in Section 3.4 which includes an overview of the entire pipeline of Bayesian HS, the details of the coarse-to-fine framework and a justification for the integration method choice made. Coarse-to-fine (CtF) Bayesian HS is thoroughly validated in Section 3.5 using both real and synthetic data. Quantitative evaluation on synthetic data reveals that even with extreme corruption by intensity noise the root-mean-square (rms) error ranges from sub-millimetre
Chapter 3. Coarse-to-fine Bayesian Helmholtz Stereopsis

Figure 3.1: Depth sampling for virtual camera pixel $p$ and its 4-connected neighbourhood.

to sub-centimetre accuracy depending on the level of input data corruption and the scale of reconstruction. The key conclusion drawn from the combined qualitative and quantitative evaluation is the superior performance, in terms of global and local shape accuracy as well as robustness to intensity noise, of the proposed Bayesian HS with the tailored integrability prior, relative to standard Maximum Likelihood Helmholtz Stereopsis (ML HS) (the state-of-the-art in BRDF independent reconstruction) and Bayesian HS with other untailored sub-optimal priors.

3.1 3D Reconstruction: problem statement and notation

Globally, the problem addressed in this work is dense surface recovery in 3D. Let us introduce the notion of a virtual camera projecting 3D points $s(x, y, z)$ onto pixel $p$ of the image plane via the projection ray $r_p$ determined by the projection function $P: p = P(s(x, y, z))$. Similarly, knowing the original 3D point’s depth $d_p$ along $r_p$ pixel $p$ can be back-projected onto $s(x, y, z)$ using the back-projection function $P_{\text{back}}: s(x, y, z) = P_{\text{back}}(p, d_p)$. This framework holds regardless of whether the projection function $P(s(x, y, z))$ of the virtual camera is perspective or orthographic. The virtual camera defines the viewpoint for the 2.5D reconstructions targeted in this work.

Further, each pixel $p$ represents a random variable $D_p$ - a measure of depth for the visible surface point projecting to $p$. The surface reconstruction problem is thus formulated as a labelling problem where depth label $d_p$ is assigned to each random variable $D_p$ of pixel $p$. For each $p$ there is a set of $N$ depth hypotheses $\{d_1, ..., d_N\}$, i.e. the possible values for $D_p$. The set of hypotheses is obtained by sampling the reconstruction volume $V$ along the pixel’s projection ray $r_p$ (Figure 3.1). The exact nature of the depth sampling depends on
the projection type of the virtual camera. In this work, the orthographic virtual camera is adopted resulting in uniform sampling of the 3D space. In this case, the framework is analogous to voxel representation where a surface is embedded in a discrete volume $V$ of $N_X \times N_Y \times N_Z$ voxels $v(x, y, z)$ sampled at resolution $\delta_x \times \delta_y \times \delta_z$. Let us define set $\mathcal{F}$ to be all the virtual camera pixels. The solution to the defined labelling problem is the label configuration $d = \{d_p | \forall p \in \mathcal{F}\}$ where $d \in S$ with $S$ being the set of all possible label configurations.

The problem lends itself well to representation as a Markov Random Field (MRF). In the MRF graph $G = (\mathcal{F}, \mathcal{E})$, each pixel $p \in \mathcal{F}$ of the virtual camera is a node. The nodes are connected by edges $e \in \mathcal{E}$ to a set of neighbouring nodes modelling spatial dependencies (Figure 3.1). These dependences define the prior probability distribution of the framework’s state variables $D_p$. Further, the composition of $\mathcal{E}$ is determined by the adopted notion of node neighbourhood. Each $D_p$ individually is also characterised by a plausibility distribution over its set of depth hypotheses $\{d_1, ..., d_N\}$, based on local observation. In practical MRF applications, the local observation gives rise to the data term whereas the prior distribution is translated into the smoothness term as many natural phenomena show a gradual evolution.

### 3.2 BRDF independent 3D Reconstruction

#### 3.2.1 Helmholtz Stereopsis

Let us now look at standard Maximum Likelihood Helmholtz Stereopsis (ML HS) in the context of adopted notation. In order to assign label $d_p$ to each pixel $p$ of the virtual camera’s image plane, a set of points $s(x, y, z)$ along the projection ray $r_p$ are sampled. Only the projection rays intersecting with the visual hull are considered. Each sampled point is defined by its position along the projection ray, which is a depth hypothesis value $d$. The depth hypothesis set $\{d_1, ..., d_N\}$ for random variable $D_p$ is thus accumulated. Let us confine the sampling range along $r_p$ to within the visual hull. The sampling resolution is arbitrary.

In HS, each $s(x, y, z)$ along $r_p$ is sampled by projection onto the reciprocal images to acquire a set of $N_w$ intensity 2-tuples $\{(i_1, i_2)_1, ..., (i_1, i_2)_N\}$ and formulate $N_w$ constraints.
The consistency measure, computed using the SVD decomposition of \( W \) and defined by the SVD-residual-based coefficient \( \sigma_2^2 \), is the local observation of \( D_p \) distributed over \( \{d_1, ..., d_N\} \). Depth hypothesis \( d_p \) of pixel \( p \) has got a likelihood \( E_{\text{data}}(p, d_p) \) defined through the SVD coefficient \( \frac{\sigma_2(P_{\text{back}}(p, d_p))}{\sigma_3(P_{\text{back}}(p, d_p))} \) associated with the corresponding 3D point \( s(x, y, z) \) that pixel \( p \) back-projects to at depth \( d_p \) along projection ray \( r_p \). The coefficient tends to infinity as \( d_p \) approaches the correct depth \( d_p^* \). For compatibility with minimisation of an MRF, in this work \( E_{\text{data}}(p, d_p) \) is formulated as a decaying function of \( \frac{\sigma_2(P_{\text{back}}(p, d_p))}{\sigma_3(P_{\text{back}}(p, d_p))} \) with the decay factor \( \mu = 0.2 \times \ln(2) \) and bounded in the range \([0, 1]\):

\[
E_{\text{data}}(p, d_p) = e^{-\mu \times \frac{\sigma_2(P_{\text{back}}(p, d_p))}{\sigma_3(P_{\text{back}}(p, d_p))}}
\]

In ML HS, depth assignment is performed based solely on this data term without involving the prior distribution from spatial dependencies. Hence, standard HS solves a non-Markovian ML optimisation problem optimising each \( D_p \) independently:

\[
d_p^*_{ML} = \arg \min_{d \in S} \sum_{p \in F} E_{\text{data}}(p, d_p)
\]

The resultant solution ignoring the prior distribution is sub-optimal. Sub-optimality leads to noisy depth maps resulting in lacking surface smoothness and structural detail. The global shape may be reasonable, but, as noisy depth labels index inconsistent normals, reconstruction finesse of ML HS is limited.

### 3.2.2 Bayesian Helmholtz Stereopsis

Instead of the sub-optimal maximum likelihood (ML) optimisation, the labelling problem of 3D reconstruction by HS is formulated in this chapter as a maximum a posteriori probability (MAP) optimisation. In this formulation, for pairs of neighbouring pixels \( p \) and
3.2. BRDF independent 3D Reconstruction

$q$, in addition to the respective data cost terms $E_{\text{data}}(p, d_p)$ and $E_{\text{data}}(q, d_q)$ as defined in Section 3.2.1, a point-to-point smoothness cost term $E_s(p, d_p, q, d_q)$ is defined. The relative data versus smoothness contribution is regulated by the normalised parameter $\alpha$. We adopt the 4-connected neighbourhood for edge set $\mathcal{E}$, computing the smoothness cost for each $(p, q) \in \mathcal{E}$. The MAP solution to the labelling problem is:

\[
d^*_{\text{MAP}} = \arg\min_{d \in S} \sum_{p \in \mathcal{F}} \left( (1 - \alpha) E_{\text{data}}(p, d_p) + \sum_{(p, q) \in \mathcal{E}} \alpha E_s(p, d_p, q, d_q) \right)
\]

(3.4)

Compared to ML HS, Bayesian formulation in (3.4) produces cleaner depth maps improving accuracy by more accurate normal indexing. A Bayesian framework is clearly more suitable because of the strong statistical dependency between neighbouring depth estimates. As the data term formulation for Bayesian HS is the same as defined in Section 3.2.1 for standard HS, in the remainder of the section we focus on the formulation and comparison of several priors to assess their relative suitability as candidates for $E_s(p, d_p, q, d_q)$. 

Figure 3.2: Schematic representation of the priors.
Depth-based prior (Dprior)

The prior schematically illustrated in Figure 3.2a is known from Conventional Stereo. The depth-based smoothness cost \( E_{s,d}(p, d_p, q, d_q) \) for neighbouring pixels \( p \) and \( q \) is defined as the discontinuity-preserving truncated squared difference of their respective depth labels \( d_p \) and \( d_q \):

\[
E_{s,d}(p, d_p, q, d_q) = \min(E_{s,d}^\text{max}, (d_p - d_q)^2)
\]

(3.5)

where the truncation value \( E_{s,d}^\text{max} \) is half the reconstruction volume squared. With penalties for large label variation between neighbouring points, the prior encourages piece-wise constant depth and biases towards a fronto-parallel representation. The effect of the prior needs to be moderated using the weighting parameter \( \alpha \) from (3.4) to some degree, depending on the depth sampling resolution, to prevent the surface going planar. Unless checked by a low \( \alpha \), at lower resolutions planarisation is unavoidable due to the large discrete cost gaps between neighbouring hypotheses. Further, in the Dprior formulation of (3.5) depth discontinuities are preserved by truncation moderating depth fluctuation penalties. Note that Dprior completely disregards the available normal information.

Normal-based prior (Nprior)

Surface characterisation through normals is typical of Photometric Stereo. A suitable normal-based prior would enforce locally constant normals encouraging locally flat, though not necessarily fronto-parallel, surfaces. Hence, Nprior is in theory less restrictive of reconstructed surfaces than the depth-based one.

Let us define function \( \mathbf{n}(p, d_p) \) as the normal estimate associated with depth hypothesis \( d_p \). Specifically, \( \mathbf{n}(p, d_p) \) is the normal vector estimate by HS at 3D point \( \mathbf{P}_{\text{back}}(p, d_p) \) - the back-projection of pixel \( p \) at depth \( d_p \) along ray \( r_p \). We define a normal-based prior where similarity of corresponding normals is used to assess neighbouring label compatibility with the discrete depth hypotheses still being the labels. Given photometric normals \( \mathbf{n}(p, d_p) \) and \( \mathbf{n}(q, d_q) \) corresponding to labels of neighbouring pixels \( p \) and \( q \) (Figure 3.2b), Nprior can be formulated as follows:

\[
E_{s,n}(p, d_p, q, d_q) = \pi^{-1} \arccos ( \mathbf{n}(p, d_p) \cdot \mathbf{n}(q, d_q) )
\]

(3.6)
The cost function in Equation 3.6 is the normalised correlation angle between normals. Complications in using normal-based priors arise because 1. normals are continuous and susceptible to noise and 2. normal correlations are irregular expressions not optimisable by graph cuts [19], one of the most commonly used MRF optimisation techniques. Consequently, sequential tree re-weighted message passing (TRW-S) [17], [18] is used for MRF optimisation consistently in this work instead because it does not require regularity of prior. Note that Nprior does not make use of the available depth information.

**Depth-normal consistency prior (DNprior)**

The depth-based prior seeks to de-noise depth maps by enforcing their smoothness, while the normal-based approach promotes gradual spatial evolution of the normal field. Both approaches are one-sided: the depth is optimised indexing the normals or vice versa. Depth and normal estimation processes are however not independent and must be consistent with each other. A superior prior in the context of HS explicitly enforces consistency between depths and normals, performing joint optimisation of the depth map and normal field. This type of smoothness term will be called the depth-normal consistency prior (DNprior).

The depth-normal consistency idea can give rise to several formulations of DNprior. The formulations are not all equally principled and any limitations directly precipitate in reconstruction artefacts. In this chapter, two novel formulations of depth-normal consistency prior are proposed: a correlation-based and a distance-based one. The superiority of the latter formulation over the former will be argued both theoretically and based on an extensive evaluation.

**Correlation-based DNprior (corr.DNprior)** is formulated as a normalised correlation angle between the geometric normal defined by the \( d_p - d_q \) depth transition from \( p \) to \( q \) and the normalised projections \( n_{\text{prj}}(p, d_p) \) and \( n_{\text{prj}}(q, d_q) \) of the corresponding estimated photometric normals onto the transition plane. In our implementation we work with an orthographic virtual camera and 4-connected pixel neighbourhoods. Figure 3.2c illustrates how the geometric normal \( n_g(p, d_p, q, d_q) \) to the \( d_p - d_q \) depth transition between back-projected points \( \mathbf{P}_{\text{back}}(p, d_p) \) and \( \mathbf{P}_{\text{back}}(q, d_q) \) is fully embedded in the transition plane because \( p \) and \( q \) are lateral neighbours in a 4-connected neighbourhood. A correct depth transition will have \( n_g(p, d_p, q, d_q) \) correlating well to photometric normals \( n_{\text{prj}}(p, d_p) \) and
\( \mathbf{n}_{\text{proj}}(q, d_q) \) given their correctness. The correlation angle quantifies the fit of labels \( d_p \) and \( d_q \) and the cost for pixel \( p \) at \( d_p \) relative to \( q \) at \( d_q \) is written as:

\[
\phi_{ph-q}(p, d_p, q, d_q) = \pi^{-1} \arccos(\mathbf{n}_{\text{proj}}(p, d_p) \cdot \mathbf{n}_g(p, d_p, q, d_q)) \tag{3.7}
\]

Since the range of the arc-cosine function is \([0, \pi]\), the orientation of \( \mathbf{n}_g \) must be forced to be consistent with the photometric normals (i.e. out of the surface). The expression \( \phi_{ph-q}(q, d_q, p, d_p) \) for pixel \( q \) at \( d_q \) relative to pixel \( p \) at \( d_p \) is analogous. Hence, the smoothness cost \( E^{\text{corr}}_{s,dn}(p, d_p, q, d_q) \) of corr.DNprior is:

\[
E^{\text{corr}}_{s,dn}(p, d_p, q, d_q) = \frac{1}{2}(\phi_{ph-q}(p, d_p, q, d_q) + \phi_{ph-q}(q, d_q, p, d_p)) \tag{3.8}
\]

**Distance-based DNprior (dist.DNprior)** is an alternative formulation of depth-normal consistency derived from the fundamental perpendicularity of the normal to the surface. Let us consider two laterally neighbouring pixels \( p \) and \( q \) from a 4-connected neighbourhood. The estimated photometric normal \( \mathbf{n}(p, d_p) \) of pixel \( p \) at depth \( d_p \) suggests a surface transition from the back-projection \( \mathbf{P}_{\text{back}}(p, d_p) \) (Figure 3.2d). Considering one direction for simplicity, mandated by \( \mathbf{n}(p, d_p) \) the surface should continue from \( \mathbf{P}_{\text{back}}(p, d_p) \) to some \( \mathbf{P}_{\text{back}}(q, d_{q_p}) \). Note that \( d_{q_p} \) does not have to be part of the discrete label set of pixel \( q \).

By definition, \( \mathbf{n}(p, d_p) \) is perpendicular to the transition plane from \( \mathbf{P}_{\text{back}}(p, d_p) \) to \( \mathbf{P}_{\text{back}}(q, d_{q_p}) \):

\[
(\mathbf{P}_{\text{back}}(q, d_{q_p}) - \mathbf{P}_{\text{back}}(p, d_p)) \cdot \mathbf{n}(p, d_p) = 0. \tag{3.9}
\]

Let us now define the depth error metric between \( d_{q_p} \) and some \( d_q \) from the set of discrete depth hypotheses of pixel \( q \). The corresponding back-projections \( \mathbf{P}_{\text{back}}(q, d_{q_p}) \) and \( \mathbf{P}_{\text{back}}(q, d_q) \) deviate only in depth:

\[
\mathbf{P}_{\text{back}}(q, d_{q_p}) = \mathbf{P}_{\text{back}}(q, d_q) + [0, 0, \delta_{q_p}]^\top \tag{3.10}
\]

where \( \delta_{q_p} \) is the depth discrepancy between \( d_q \) and \( d_{q_p} \). The smaller the \( \delta_{q_p} \), the greater the confidence of \( d_p \) and \( d_q \) being the right labels for pixels \( p \) and \( q \) respectively. Substituting
(3.10) into (3.9) we get:

\[
(P_{\text{back}}(q, d_q) + [0, 0, \delta_{qp}]^T - P_{\text{back}}(p, d_p)) \cdot n(p, d_p) = 0.
\]  

(3.11)

Since \(\delta_{qp}\) is aligned with the depth dimension, along ray \(r_p\), its dot product is only dependent on that dimension of the normal reducing (3.11) to:

\[
(P_{\text{back}}(q, d_q) - P_{\text{back}}(p, d_p)) \cdot n(p, d_p) + \delta_{qp} n_{r_p}(p, d_p) = 0
\]  

(3.12)

where \(n_{r_p}(p, d_p)\) is the depth component of the photometric normal along projection ray \(r_p\). Finally, the following expression unambiguously defines depth discrepancy \(\delta_{qp}\) between \(d_q\) and \(d_{qp}\) (Figure 3.2d):

\[
\delta_{qp} = \frac{(P_{\text{back}}(q, d_q) - P_{\text{back}}(p, d_p)) \cdot n(p, d_p)}{n_{r_p}(p, d_p)}.
\]  

(3.13)

Hence the proposed distance-based formulation of the depth-normal consistency prior minimises the depth discrepancy of (3.13). Depth discrepancy is squared to prevent the staircase error common with depth-based priors where the cost of a set of small depth steps exceeds that of a single large step. Every edge \((p, q) \in \mathcal{E}\) is characterised by the symmetrical pair of depth discrepancy scores, \(\delta_{qp}\) and \(\delta_{pq}\), each contributing equally to the final smoothness cost. The final cost \(E_{\text{dist}, \text{dn}}^{\text{dist}}(p, d_p, q, d_q)\) of \(\text{dist.DNprior}\) reads:

\[
E_{\text{dist}, \text{dn}}^{\text{dist}}(p, d_p, q, d_q) = \frac{1}{2}(\delta_{qp})^2 + (\delta_{pq})^2.
\]  

(3.14)

The physical meaning of \(\text{dist.DNprior}\) is intuitively easier to grasp than the correlations of \(\text{corr.DNprior}\). The distance-based formulation is plausible being derived from the surface normal axiom and is consistent with the formulation of the well-established depth prior. The following section builds an argument providing theoretical grounds to support intuition in showing the distance-based \(\text{DNprior}\) formulation to be more principled than the correlation-based one. The argument rests on the relation of \(\text{dist.DNprior}\) to integrability as a universally desirable surface property.
3.3 Prior optimality

Optimality of dist.DNprior can be argued on the grounds of its smoothness function $E_{\text{DN}}(p, d_p, q, d_q)$ being equivalent to the surface integrability constraint. Integrability is the least restrictive prior. Unlike priors biasing towards fronto-parallel (Dprior) or locally flat (Nprior) surfaces, the integrability constraint is essentially just a check of mathematical plausibility and will only definitively bias against extremes such as Dirac peaks (e.g. the surface of a cactus) or sharp step-like transitions. As such drastic depth fluctuations are rare, the integrability constraint is most widely applicable.

The integrability constraint forms the *height-from-gradient* recovery framework from the classical paper by Horn [111]. In [111], surface gradient $(g_x, g_y)$ is the derivative of surface $z(x, y)$: $g_x = \frac{\delta z}{\delta x}$ and $g_y = \frac{\delta z}{\delta y}$. The surface normal is related to the gradient as $\mathbf{n} = (1 + g_x^2 + g_y^2)^{-\frac{1}{2}}(-g_x, -g_y, 1)^T$. The surface gradient can be related to the unit surface normal via: $g_x = -\frac{n_x}{n_z}$ and $g_y = -\frac{n_y}{n_z}$. Hence the following cost function formulated by Horn is a joint optimisation of surface height (or equivalently depth) $z(x, y)$ and normal over the entire surface:

$$\int \int ((\frac{\delta z}{\delta x} - g_x)^2 + (\frac{\delta z}{\delta y} - g_y)^2) \, dx \, dy$$

Clearly, the cost function encapsulates the idea of depth-normal consistency of DNprior. For clarity, let us consider surface evolution along the $x$ dimension only. The $y$ dimension is analogous. Numerical approximation of gradient discretises the derivative to: $\frac{\delta z}{\delta x} = \frac{z_2 - z_1}{\delta x}$ where $z_1 = z(x, y)$ and $z_2 = z(x + \delta x, y)$. In the terminology of priors presented earlier $z_1$ and $z_2$ are depth labels $d_p$ and $d_q$ of two neighbouring pixels $p$ and $q$. Using gradient discretisation the integrability prior function $E_{\text{DN}}^{Horn}$ based on (3.15), symmetrical with respect to neighbouring points, is formulated as follows:

$$E_{\text{dn}}^{Horn} = \frac{1}{2} \left( (\frac{z_2 - z_1}{\delta x} - g_{x,1})^2 + (\frac{z_2 - z_1}{\delta x} - g_{x,2})^2 \right) = \frac{1}{2} e_{1,\text{dn}}^{Horn} + \frac{1}{2} e_{2,\text{dn}}^{Horn}$$

where $g_{x,1}$ and $g_{x,2}$ are the $x$ gradients of two neighbouring 3D points $s_1(x, y)$ and $s_2(x + \delta x, y)$ expressed in terms of unit normal components.

Let us now relate dist.DNprior to the optimal integrability prior $E_{\text{dn}}^{Horn}$. Depths $d_{qp}$ and
3.3. Prior optimality

d_{pq} become \(z_{21}\) and \(z_{12}\) i.e. the neighbouring depth labels suggested by the normal-imposed
gradients: \(z_{21} = z_1 + \delta x g_{x,1}\) and \(z_{12} = z_2 - \delta x g_{x,2}\). Substituting the definitions, the
corresponding depth discrepancies \(\delta_{q_p} = \delta_{21}\) and \(\delta_{p_q} = \delta_{12}\) are: \(\delta_{21} = (z_{21} - z_2) = (z_1 + \delta x g_{x,1} - z_2)\) and \(\delta_{12} = (z_{12} - z_1) = (z_2 - \delta x g_{x,2} - z_1)\). Assuming two given neighbouring
3D points, the distance-based DNprior cost from (3.14) becomes:

\[
E_{dn}^{dist} = \frac{1}{2} ((z_1 + \delta x g_{x,1} - z_2)^2 + (z_2 - \delta x g_{x,2} - z_1)^2) \tag{3.17}
\]

The equation can be manipulated into the following:

\[
E_{dn}^{dist} = \frac{1}{2} \left( \frac{z_2 - z_1}{\delta x} - g_{x,1} \right)^2 + \left( \frac{z_2 - z_1}{\delta x} - g_{x,2} \right)^2 \right) \delta x^2
\]

\[
= \left( \frac{1}{2} e_{1,dn}^{Horn} + \frac{1}{2} e_{2,dn}^{Horn} \right) \delta x^2 = E_{dn}^{Horn} \delta x^2 \tag{3.18}
\]

Equation (3.18) shows equivalence of dist.DNprior to the surface integrability prior \(E_{dn}^{Horn}\) at resolution \(\delta x\).

Despite also being based on depth-normal consistency, corr.DNprior does not reduce directly
to the integrability prior. We state only the key results leaving the derivation details to the
appendix. To facilitate comparison to the surface integrability prior \(E_{dn}^{Horn}\), the correlation
angle of \(E_{sdn}^{corr}\) is re-formulated in terms of the normalised tangent \(t_x\) to the depth transition
and the normalised normal projection vector: \(\phi_{ph-g} = \pi^{-1} |arcsin(n_{prj,x} \cdot t_x)|\) which is
equivalent to (3.7). The square of the dot product reduces to:

\[
(n_{prj,1,x} \cdot t_x)^2 = \frac{e_{1,dn}^{Horn}}{e_{1,dn}^{Horn} + (\frac{z_2-z_1}{\delta x} - g_{x,1} + 1)^2} \tag{3.19}
\]

It has been seen that dist.DNprior in (3.18) boils down to Horn’s surface integrability
in (3.16) at a given resolution. No such equivalence of corr.DNprior to integrability can
be concluded by substitution of (3.19) into the \(\phi_{ph-g}\) expression. Hence, as postulated,
the distance-based depth-normal consistency prior is a more principled formulation being,
unlike the correlation-based one, directly equivalent at a given resolution to the surface
integrability prior, the optimal least restrictive regularisation principle.

It has been observed in literature [111], [64] that the non-restrictive integrability prior,
enforcing nothing more than the mathematical surface plausibility, may be insufficient for
correct reconstruction in under-constrained optimisation schemes with a weak data term (e.g. Shape from Shading, 2-constraint Photometric Stereo). As Horn originally stated, starting the system from distant estimates requires more restrictive regularisation terms in addition to integrability. In the current work with a sufficient ($\geq 3$) number of HS constraints, the local observation distribution is reliable enough to resolve remaining ambiguity with the integrability prior alone permitting most unconstrained geometries and avoiding any ungrounded biases.

### 3.4 Implementation

**The overall pipeline for Bayesian HS.** An overview of the Bayesian HS reconstruction pipeline is presented in Figure 3.3. A set of reciprocal image pairs with the minimum cardinality of three is acquired and used to compute constraints in Equation 3.1 for each voxel in the swept reconstruction volume. The matrix of constraints is subjected to SVD decomposition to characterise each voxel by a normal in addition to its depth attribute. Using this joint depth and normal information the smoothness term is computed per voxel while the data term is defined by the SVD residual signifying the quality of normal estimate. The subsequent MRF optimisation using TRW-S [17], [18] with six passes solves the Bayesian problem seeking the best set of voxels in the volume to represent the surface. In order to refine the surface the reconstruction sequence is repeated several times in a coarse-to-fine framework described in more detail further on in this section.

Depending on the resolution of the generated point cloud (i.e. the number of coarse-to-fine iterations computed), the integration can be performed with either Poisson Surface Reconstruction (PR) for low resolution and by ordering vertices into facets based on known geometric relationships within the reconstruction volume (NoInt) for high resolution. A discussion towards justification of the integration method choice is also given further on in the section. Since the point clouds of high densities are showcased in this chapter to demonstrate the full potential of the reconstruction pipeline, NoInt is mostly used (except in Section 3.5.2 where the performance of three integration methods is evaluated relative to each other).

**Coarse-to-fine reconstruction.** Bayesian HS with a suitable prior allows previously unattainable reconstruction accuracy. The accuracy is however always fundamentally lim-
3.4. Implementation

The resolution of the original pixel grid is doubled by subdividing each pixel into 4 superpixels. The depth sampling resolution within $\delta d_{p,0}$ is doubled as well with two new hypotheses being introduced to replace each initial depth label. The superpixels inherit the reduced depth search space of the parent pixel but with refined sampling $\frac{\delta z_0}{2}$ and are subsequently labelled out of their individual search spaces. For an orthographic camera, the sub-division of the pixel grid and the depth label space together is analogous to refining resolution of the voxel volume with every original voxel giving rise to eight sub-voxels. Spatial sub-division is stopped, while the depth label resolution doubling persists, when neighbouring sub-voxels start projecting onto the same pixel. The final output of the process is a point cloud of vertices $P_{back}(p, d_{p,L_{max}})$ each with a normal estimate.
\( \mathbf{n}(p, d_{p,L_{\max}}) \) where \( p \) is a superpixel in the final optimisation pass \( L_{\max} \) and \( d_{p,L_{\max}} \) is the corresponding depth label.

**Integration method.** The resolution boost via the CtF Bayesian HS framework with integrability prior allows the unprecedented in HS explicit integration elimination. In standard ML HS, a one-sided approach of integrating normal fields using the Fourier-based Frankot-Chellappa algorithm (FC) is employed. As fusion of depth and normal estimates is beneficial at both estimation and integration stages, Poisson Surface Reconstruction (PR) on the point cloud of oriented vertices generated by the reconstruction algorithm is a more holistic method taking advantage of the uniquely dual depth/normal vertex characterisation by HS. PR outperforms FC by avoiding global drift and integrability enforcing approximations of direct normal integration. Despite being competitive, PR is an additional post-processing optimisation procedure that will alter the point cloud estimated at reconstruction. Hence, for reasons of maximal consistency with the input point cloud and for computational efficiency, an integration-free Bayesian HS pipeline with integrability prior is advocated in this work. In the pipeline finely sampled vertices of dense point clouds are arranged into continuous geometries using spatial proximity determined by the vertices belonging to 4-connected neighbourhoods (Figure 3.1). Further, either the geometric or photometric normals define the continuous surface shading. The geometric normals are determined by facet orientation of the computed geometry (hence derived from the HS depth estimates) and give rise to the so-called flat shading. The photometric normals are the per-vertex estimates by HS defining the smooth shading of the reconstructed surface.

The elimination of post-processing with complex integration algorithms such as FC and PR...
3.5. Evaluation

is not merely an efficiency improvement in dense point cloud reconstruction by Bayesian HS, but primarily a safeguard against input point cloud alterations by drift or oversmoothing. As has been mentioned however for less dense point clouds computed with a lower iteration count of the CtF framework, PR is still as excellent choice for integration method as NoInt in this case will produce results too faceted in appearance. Finally, FC is only warranted in the case of point clouds with extremely low depth estimate quality such as those of standard ML HS where normals are the main carriers of geometry related information. The shape integrated by FC at poor depth locations will not be correct but it will be recognisable as FC enforces integrability which is the best one can hope for without accurate depth information. For illustration the reader is referred to the appendix where the surfaces by the three integration methods on ML HS point clouds are presented for a variety of datasets.

3.5 Evaluation

The evaluation targets verification of several claims quantitatively and qualitatively on versatile synthetic (Section 3.5.1) and real (Section 3.5.2) data. Discussed are the benefit of the proposed Bayesian HS formulation relative to its standard ML one and subsequently, the superior performance of the distance-based DNprior over Dprior, Nprior and the correlation-based DNprior within the Bayesian formulation. Quantitative evaluation in Section 3.5.1 incorporates a numerical comparison of the different methods based on accuracy, completeness and robustness to varying levels of intensity noise. In Section 3.5.2, the advantage of the proposed coarse-to-fine approach is shown which facilitates a pipeline without explicit surface integration. Compared are the final meshes integrated using 1. the Fourier-based Frankot-Chellappa algorithm (FC); 2. Poisson Surface Reconstruction (PR) and 3. the explicit-integration-free approach ordering vertices into facets based on known geometric relationships within the reconstruction volume (NoInt).

3.5.1 Synthetic Data

Using synthetic data we quantify reconstruction accuracy and completeness and show the robustness of the proposed pipeline against simulated intensity noise. The three synthetic datasets rendered with Povray [112] are a simple sphere (“Sphere”), the pear from [113]
Chapter 3. Coarse-to-fine Bayesian Helmholtz Stereopsis

Figure 3.5: Examples of input intensity images (16-bit, grayscale) for three synthetic datasets at three intensity noise levels.

(“Pear”) and the Stanford bunny [114] (“Bunny”) in Figure 3.5. Each dataset comprises eight reciprocal pairs at three intensity noise levels: noise free and noise levels 1 and 2 with the normalised variances of 0.001 and 0.01 (±2072 and ±6553 intensity levels) respectively. The noise corruption is extreme to fully test robustness. The synthetic objects are reconstructed using standard HS (1) and Bayesian HS with Nprior (2), Dprior (3), corr.DNprior (4) and dist.DNprior (5). The reconstruction is coarse-to-fine with five iterations and NoInt is used to obtain the final surface.

Performance is measured using accuracy and completeness from [24]: the accuracy of $x \, mm$ at threshold $\theta = N\%$ indicates that $N\%$ of reconstruction vertices lie within $x \, mm$ of the ground truth whereas the completeness of $N\%$ at threshold of $x \, mm$ means that, when the inlier threshold is set at $x \, mm$, only $N\%$ of the ground truth vertices are estimated.

Scale dependency of the distance-based metrics results in different smoothness weights of Dprior $\alpha_d$ between the datasets. The same dependency affects dist.DNprior due to $\delta x$ in (3.18) necessitating an additional scale-related prior cost adjustment factor $\beta_{dist, dn}$ deriv-
able from the relative spatial reconstruction resolution: since at \(1 \text{ mm}\) resolution no scale adjusting proves necessary, at \(10 \text{ mm}\) resolution to obtain good results the \(\delta x\) to \(10\delta x\) scale adjustment in (3.18) is required (the corresponding cost adjustment factor \(\beta_{\text{dist,\,dn}}\) is 100).

The MRF cost component becomes: 

\[(1 - \alpha)E_{\text{data}}(p, d_p) + \beta_{\text{dist,\,dn}}\alpha E_{s}(p, d_p, q, d_q)\]

Further, in reconstructing the bunny, the dist.DNprior cost is truncated at \(2.9 \text{ mm}\) in the first iteration, with the value decreasing with the reconstruction volume in the subsequent iterations. Truncation, only relevant for the intricate bunny, prevents high-frequency geometry, that is ray-traced to a limited precision only, from giving rise to global shape distortions.

The reconstruction volume sizes \(|V|\), initial sampling (spatial \(\delta x/\delta y\)/depthwise \(\delta z\), the smoothness weighting \(\alpha\) (per prior i.e. \(\alpha_n, \alpha_d, \alpha_{\text{corr,\,dn}}\) and \(\alpha_{\text{dist,\,dn}}\)) and the adjustment factors \(\beta_{\text{dist,\,dn}}\) (per dataset) are as stated in Figures 3.6, 3.8 and 3.10.

The table in Figure 3.12 presents the Middlebury metrics for the three datasets at accuracy thresholds \(th\) of 70\%, 90\% and 100\% with the top performer per dataset and noise level indicated in bold. Graphs of the performance metrics within the full \(th\). sweep can be found in Figures 3.7, 3.9 and 3.11 for Sphere, Pear and Bunny respectively. Dprior, Nprior and standard HS are not in the same league as corr.DNprior and dist.DNprior with their accuracy error scores mostly an order of magnitude higher. Similarly, even with an unrestricted inlier threshold these methods are not always capable of matching the completeness score of the two better priors. Dist.DNprior shows a better robustness consistently outperforming corr.DNprior on the noise corrupted data. For noiseless data, dist. and corr.DNprior mostly perform quite similarly (e.g. Sphere and Bunny) with dist.DNprior only clearly dominating on Pear. The scores for Bunny at 100\% are not indicative of the overall performance being heavily influenced by few strong edge-based outliers. The final completeness scores of corr.DNprior and dist.DNprior are similar (72\%, 93\% and 90\% for Sphere, Pear and Bunny respectively) but at different inlier thresholds.

The error maps in Figures 3.6, 3.8 and 3.10 spatially rather than statistically compare the reconstructions to the ground truth at different levels of corruption by intensity noise. The drastic scale variation between the error maps is indicative of the methods’ relative performance (both corr.DNprior and dist.DNprior are represented on the same scale). Consistent with its lack of regularisation standard HS produces noisy results whereas Dprior and Nprior generate either extreme global deformations or random noise. Neither dist.DNprior nor
corr.DNprior is susceptible to random noise as much as the inferior priors but corr.DNprior does show shape distortions, both local and global, when reconstructing from noise corrupted data. In particular, note the typical step-like artefact observed by inspection of the flat-shaded corr.DNprior reconstructions which becomes pronounced with increasing intensity noise. In contrast, both the flat-shaded meshes and the corresponding error heat maps confirm the greater resilience of dist.DNprior to both global and local shape distortion in the face of random intensity noise.

While showing the same inter-prior trends, the errors of Sphere exceed those of Pear and Bunny because the former, being larger, is reconstructed at the cm, rather than mm, scale. For heavily corrupted data (noise level 2), at 90% all vertices lie within 16.66 mm (Sphere), 1.90 mm (Pear) and 3.29 mm (Bunny) when using Bayesian HS with dist.DNprior. The result is not bad considering that the level of input data corruption is extreme by far exceeding the expectation in real-life acquisition scenarios. The corresponding measurements with noiseless data are the exceptionally good nearly universally sub-millimetre values of 1.05 mm (Sphere), 0.27 mm (Pear) and 0.81 mm (Bunny) which is an order of magnitude better than at noise level 2.
3.5. Evaluation

<table>
<thead>
<tr>
<th></th>
<th>Flat-shaded reconstruction</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>standard HS</td>
<td>Nprior</td>
</tr>
<tr>
<td>no noise</td>
<td><img src="image1" alt="Flat-shaded reconstruction" /></td>
<td><img src="image2" alt="Nprior" /></td>
</tr>
<tr>
<td>noise level 1</td>
<td><img src="image1" alt="Flat-shaded reconstruction" /></td>
<td><img src="image2" alt="Nprior" /></td>
</tr>
<tr>
<td>noise level 2</td>
<td><img src="image1" alt="Flat-shaded reconstruction" /></td>
<td><img src="image2" alt="Nprior" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Error heat maps</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>standard HS</td>
<td>Nprior</td>
</tr>
<tr>
<td>no noise</td>
<td><img src="image6" alt="Error heat maps" /></td>
<td><img src="image7" alt="Nprior" /></td>
</tr>
<tr>
<td>noise level 1</td>
<td><img src="image6" alt="Error heat maps" /></td>
<td><img src="image7" alt="Nprior" /></td>
</tr>
<tr>
<td>noise level 2</td>
<td><img src="image6" alt="Error heat maps" /></td>
<td><img src="image7" alt="Nprior" /></td>
</tr>
</tbody>
</table>

Figure 3.6: Synthetic Sphere. Parameters: $|V| = 400\text{mm} \times 400\text{mm} \times 600\text{mm}$, initial resolution: 10 mm/1 mm (spatially/depthwise); smoothness weight $\alpha$ per prior: $\alpha_n = 0.5$, $\alpha_d = 10^{-7}$, $\alpha_{corr,dn} = 0.8$ and $\alpha_{dist,dn} = 0.8$; cost adjustment to scale: $\beta_{dist,dn} = 1$. 
<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Completeness</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
</tr>
<tr>
<td><img src="image3" alt="Graph" /></td>
<td><img src="image4" alt="Graph" /></td>
</tr>
<tr>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
</tr>
<tr>
<td><img src="image7" alt="Graph" /></td>
<td><img src="image8" alt="Graph" /></td>
</tr>
</tbody>
</table>

Figure 3.7: Synthetic Sphere (continued): Middlebury accuracy and completeness graphs for Bayesian HS with different priors and ML HS.
### Flat-shaded reconstruction

<table>
<thead>
<tr>
<th></th>
<th>standard HS</th>
<th>Nprior</th>
<th>Dprior</th>
<th>corr.DNprior</th>
<th>dist.DNprior</th>
</tr>
</thead>
<tbody>
<tr>
<td>no noise</td>
<td><img src="image1.png" alt="image" /></td>
<td><img src="image2.png" alt="image" /></td>
<td><img src="image3.png" alt="image" /></td>
<td><img src="image4.png" alt="image" /></td>
<td><img src="image5.png" alt="image" /></td>
</tr>
<tr>
<td>noise level 1</td>
<td><img src="image6.png" alt="image" /></td>
<td><img src="image7.png" alt="image" /></td>
<td><img src="image8.png" alt="image" /></td>
<td><img src="image9.png" alt="image" /></td>
<td><img src="image10.png" alt="image" /></td>
</tr>
<tr>
<td>noise level 2</td>
<td><img src="image11.png" alt="image" /></td>
<td><img src="image12.png" alt="image" /></td>
<td><img src="image13.png" alt="image" /></td>
<td><img src="image14.png" alt="image" /></td>
<td><img src="image15.png" alt="image" /></td>
</tr>
</tbody>
</table>

### Error heat maps

<table>
<thead>
<tr>
<th></th>
<th>standard HS</th>
<th>Nprior</th>
<th>Dprior</th>
<th>corr.DNprior</th>
<th>dist.DNprior</th>
</tr>
</thead>
<tbody>
<tr>
<td>no noise</td>
<td><img src="image16.png" alt="image" /></td>
<td><img src="image17.png" alt="image" /></td>
<td><img src="image18.png" alt="image" /></td>
<td><img src="image19.png" alt="image" /></td>
<td><img src="image20.png" alt="image" /></td>
</tr>
<tr>
<td>noise level 1</td>
<td><img src="image21.png" alt="image" /></td>
<td><img src="image22.png" alt="image" /></td>
<td><img src="image23.png" alt="image" /></td>
<td><img src="image24.png" alt="image" /></td>
<td><img src="image25.png" alt="image" /></td>
</tr>
<tr>
<td>noise level 2</td>
<td><img src="image26.png" alt="image" /></td>
<td><img src="image27.png" alt="image" /></td>
<td><img src="image28.png" alt="image" /></td>
<td><img src="image29.png" alt="image" /></td>
<td><img src="image30.png" alt="image" /></td>
</tr>
</tbody>
</table>

Figure 3.8: Synthetic Pear: Reconstructions and depth error heat maps. Parameters: $|V|=110mm \times 140mm \times 100mm$, initial resolution: $1mm/0.25mm$ (spatially/depthwise); smoothness weight $\alpha$ per prior: $\alpha_n = 0.5$, $\alpha_d = 0.05$, $\alpha_{corr, dn} = 0.8$ and $\alpha_{dist, dn} = 0.8$; cost adjustment to scale: $\beta_{dist, dn} = 100$. 
Figure 3.9: Synthetic Pear (continued): Middlebury accuracy and completeness graphs for Bayesian HS with different priors and ML HS.
### Flat-shaded reconstruction

<table>
<thead>
<tr>
<th></th>
<th>standard HS</th>
<th>Nprior</th>
<th>Dprior</th>
<th>corr.DNprior</th>
<th>dist.DNprior</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>no noise</strong></td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
<tr>
<td><strong>noise level 1</strong></td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
<tr>
<td><strong>noise level 2</strong></td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
</tbody>
</table>

### Error heat maps

<table>
<thead>
<tr>
<th></th>
<th>standard HS</th>
<th>Nprior</th>
<th>Dprior</th>
<th>corr.DNprior</th>
<th>dist.DNprior</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>no noise</strong></td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
<tr>
<td><strong>noise level 1</strong></td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
<tr>
<td><strong>noise level 2</strong></td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
</tbody>
</table>

Figure 3.10: Synthetic Stanford Bunny. Parameters: $|V| = 160mm \times 270mm \times 100mm$, initial resolution: 1 mm/0.25 mm (spatially/depthwise); smoothness weight $\alpha$ per prior: $\alpha_n = 0.5$, $\alpha_d = 0.05$, $\alpha_{corr,dn} = 0.8$ and $\alpha_{dist,dn} = 0.8$; cost adjustment to scale: $\beta_{dist,dn} = 100$. 
Figure 3.11: Synthetic Stanford Bunny (continued): Middlebury accuracy and completeness graphs for Bayesian HS with different priors and ML HS.
### Evaluation

<table>
<thead>
<tr>
<th>th.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sphere</td>
<td>9.83</td>
<td>41.71</td>
<td>35.50</td>
<td>0.53</td>
<td>0.56</td>
<td>32.62</td>
<td>26.08</td>
<td>23.40</td>
<td>5.10</td>
<td>3.32</td>
<td>87.40</td>
<td>51.16</td>
<td>61.98</td>
<td>19.73</td>
<td>15.85</td>
</tr>
<tr>
<td>70%</td>
<td>22.63</td>
<td>50.48</td>
<td>39.60</td>
<td>1.00</td>
<td>1.05</td>
<td>77.25</td>
<td>35.24</td>
<td>35.65</td>
<td>6.37</td>
<td>4.05</td>
<td>202.28</td>
<td>74.09</td>
<td>108.66</td>
<td>22.22</td>
<td>16.66</td>
</tr>
<tr>
<td>90%</td>
<td>494.46</td>
<td>88.22</td>
<td>42.67</td>
<td>2.76</td>
<td>2.76</td>
<td>451.46</td>
<td>156.64</td>
<td>104.08</td>
<td>8.76</td>
<td>5.55</td>
<td>473.78</td>
<td>148.55</td>
<td>323.82</td>
<td>26.95</td>
<td>18.28</td>
</tr>
<tr>
<td>100%</td>
<td>72.1%</td>
<td>72.1%</td>
<td>72.1%</td>
<td>72.1%</td>
<td>71.8%</td>
<td>71.8%</td>
<td>71.8%</td>
<td>72.0%</td>
<td>72.1%</td>
<td>69.0%</td>
<td>70.5%</td>
<td>69.7%</td>
<td>71.8%</td>
<td>72.0%</td>
<td>65.5%</td>
</tr>
<tr>
<td>Pear</td>
<td>6.84</td>
<td>19.53</td>
<td>14.75</td>
<td>0.22</td>
<td>0.14</td>
<td>21.44</td>
<td>25.71</td>
<td>15.68</td>
<td>0.57</td>
<td>0.37</td>
<td>33.91</td>
<td>39.85</td>
<td>21.50</td>
<td>1.90</td>
<td>1.25</td>
</tr>
<tr>
<td>70%</td>
<td>13.62</td>
<td>26.26</td>
<td>16.38</td>
<td>0.40</td>
<td>0.27</td>
<td>42.33</td>
<td>36.18</td>
<td>17.74</td>
<td>0.95</td>
<td>0.62</td>
<td>54.48</td>
<td>55.58</td>
<td>24.55</td>
<td>2.59</td>
<td>1.90</td>
</tr>
<tr>
<td>90%</td>
<td>59.30</td>
<td>55.22</td>
<td>17.94</td>
<td>1.77</td>
<td>1.47</td>
<td>82.45</td>
<td>58.01</td>
<td>21.73</td>
<td>2.08</td>
<td>1.93</td>
<td>82.97</td>
<td>83.41</td>
<td>29.83</td>
<td>4.11</td>
<td>2.80</td>
</tr>
<tr>
<td>100%</td>
<td>93.5%</td>
<td>93.6%</td>
<td>93.7%</td>
<td>93.5%</td>
<td>92.9%</td>
<td>93.4%</td>
<td>93.6%</td>
<td>93.7%</td>
<td>93.7%</td>
<td>90.0%</td>
<td>89.8%</td>
<td>93.4%</td>
<td>93.7%</td>
<td>93.7%</td>
<td>63.1%</td>
</tr>
<tr>
<td>Bunny</td>
<td>3.83</td>
<td>2.95</td>
<td>4.03</td>
<td>0.47</td>
<td>0.38</td>
<td>10.72</td>
<td>10.29</td>
<td>7.30</td>
<td>1.21</td>
<td>0.42</td>
<td>26.85</td>
<td>29.81</td>
<td>15.32</td>
<td>3.78</td>
<td>1.73</td>
</tr>
<tr>
<td>70%</td>
<td>6.89</td>
<td>5.67</td>
<td>7.64</td>
<td>0.85</td>
<td>0.81</td>
<td>20.85</td>
<td>19.75</td>
<td>13.36</td>
<td>2.26</td>
<td>0.91</td>
<td>54.31</td>
<td>56.19</td>
<td>23.58</td>
<td>5.91</td>
<td>3.29</td>
</tr>
<tr>
<td>90%</td>
<td>74.00</td>
<td>55.18</td>
<td>44.62</td>
<td>49.83</td>
<td>70.86</td>
<td>85.72</td>
<td>72.78</td>
<td>45.91</td>
<td>50.52</td>
<td>41.28</td>
<td>90.61</td>
<td>96.97</td>
<td>47.98</td>
<td>51.67</td>
<td>76.86</td>
</tr>
<tr>
<td>100%</td>
<td>90.2%</td>
<td>90.3%</td>
<td>90.5%</td>
<td>89.7%</td>
<td>90.3%</td>
<td>89.1%</td>
<td>89.4%</td>
<td>90.5%</td>
<td>90.5%</td>
<td>87.1%</td>
<td>87.6%</td>
<td>90.3%</td>
<td>89.9%</td>
<td>90.6%</td>
<td>64.9%</td>
</tr>
</tbody>
</table>

Figure 3.12: Middlebury quantitative quality assessment metrics for synthetic data: \( \text{th.} \% \) - Middlebury accuracy threshold; \( \text{meth.} \) - reconstruction method: 1. standard HS and Bayesian HS with 2. Nprior, 3. Dprior, 4. corr.DNprior and 5. dist.DNprior.
Chapter 3. Coarse-to-fine Bayesian Helmholtz Stereopsis

3.5.2 Real Data

The real data adopted from [110] consist of some clearly specular objects (teapots, vase, billiard ball), semi-specular mannequin and cup, a Lambertian terracotta doll and the highly textured teddy-bear (Figure 3.13). The geometric complexity ranges from straightforward spherical to the intricate doll. Each dataset contains eight reciprocal pairs. The datasets are processed at the mm (spatial/depthwise) initial resolution (teddy bear: 1 mm/0.5 mm; the rest: (1 mm/0.25 mm)) and the reconstruction volume sizes $|V|$ are as indicated per dataset in Figure 3.14. As for synthetic data, the reconstruction is performed using standard ML HS (1) and Bayesian HS with Nprior (2), Dprior (3), corr.DNprior (4) and dist.DNprior (5) and integrated after the 5th iteration of the coarse-to-fine framework. The weighting $\alpha$ for the priors are: $\alpha_n = 0.5$ (billiard ball: $\alpha_n = 0.1$), $\alpha_d = 0.05$ (billiard ball: $\alpha_d = 0.003$) and $\alpha_{corr,dn}, \alpha_{dist,dn} = 0.8$ without further prior adjustment or truncation. The integration is performed by ordering vertices into facets based on known geometric relationships within the reconstruction volume (NoInt) in Figure 3.14. A comparison of NoInt to the Frankot-Chellappa algorithm (FC) and Poisson Surface Reconstruction (PR) as integration methods on the best reconstruction results is presented in Figure 3.19 and discussed further on.

In total 15 (5 reconstruction $\times$ 3 integration methods) pipelines are compared per dataset.
The complete set of results (including traditional depth maps, front and side views for different integration methods etc.) are presented in the appendix. In this chapter a sub-set with the most representative results and key result representations is highlighted. Specifically, in Figure 3.14 flat-shaded meshes are shown for accuracy assessment of pure geometry and its comparison between the five reconstruction methods. The flat-shaded meshes can be viewed as 3D depth maps (conventional 2D maps can be found in the appendix.) Further, to compare the accuracy of obtained photometric information in the form of estimated vertex normals, the corresponding smooth-shaded meshes are given in Figure 3.15. Conventional RGB normal maps are also provided in Figure 3.16. Some geometric distortions hidden in the frontal view are more apparent in the side views in Figure 3.16. The relative performance of the five reconstruction methods based on the presented results is discussed next.

**Reconstruction method**

Figures 3.14 and 3.15 compare the reconstruction methods based on the final surfaces obtained by NoInt (direct meshing) with flat and smooth shading respectively. A flat-shaded surface is the best representation for structural accuracy analysis and the relative trends revealed are consistent with those observed on synthetic data. Standard HS and Bayesian HS with Nprior are quite similar in behaviour being characterised primarily by excessive noise. Nprior fails to regularise HS because well-correlating normals do not necessarily index vertices forming a plausible shape, globally or locally i.e. depth jumps are not penalised if the normals correlate well. While improving the local shape and reducing noise, Dprior, also being a limited one-sided prior, produces a typical specularity-induced shape distortion (e.g. the teapots and billiard ball). Corr.DNprior and dist.DNprior optimise depth and normals jointly, which leads to a superior performance both in terms of global shape and resolution. The performance of corr.DNprior is however consistently plagued by the abrupt step-like depth transition artefact also seen on the synthetic data. Dist.DNprior, on the other hand, inherently enforcing integrability, is not prone to forming such depth plateaus and reveals structurally accurate high-resolution meshes in the flat-shaded, smooth-shaded and normal map representations.

Smooth-shaded meshes hide all manner of sin in the reconstruction with photometric normal
estimates creating an illusion of a recognisable object even with the noisiest depth maps (e.g. standard ML HS and Nprior). However the side views of the smooth-shaded meshes in Figure 3.17 reveal drastic global shape distortions for even the best of the inferior priors (Dprior). The global shape of the superior DNpriors is in contrast very plausible in the side view as well as the previous discussed front view. The side views once again exhibit the typical step-like artefact of corr.DNprior. The artefact is also evident in the normal maps in Figure 3.16 from the distinct colour zones in the RGB representation (e.g. teapot nr. 2, mannequin, doll). The proposed dist.DNprior is completely free of such artefacts.

Figure 3.18 shows how intensity sampling with patch-based averaging inspired by [110] and adapted in this work for a coarse-to-fine framework can be used to lessen pseudo-geometries i.e. texture wrongly reconstructed as part of geometry. The artefact can be seen on the teapots, vase and billiard ball in all the representations above. In the case of the teapots and the vase, the interfering texture is the imprinted pattern on the porcelain whereas for the billiard ball it is the specularity with sensor saturation. The pseudo-geometry artefact occurs because intensities within a single reciprocal pair are sampled from different intensity fields resulting in a normal indicating a geometric slope, further sharpened in the later iterations of the coarse-to-fine framework. This misprojection occurs even for the hypotheses close to the true depth. As a remedy, for the results in Figure 3.18, intensity is sampled as the average over an image patch whose size is defined by both texture feature size and sampling resolution. The pseudo-geometries are substantially mitigated as the result. In the appendix for each dataset affected by pseudo-geometries along with the standard pointwise sampling results the reader can find a complete set of all 15 reconstructions (5 reconstruction methods × 3 integration methods) obtained with patch-based averaging for intensity sampling.
Figure 3.14: Reconstruction of real static datasets (frontal view). Reconstruction cores compared: 1. standard HS and Bayesian HS with 2. Nprior, 3. Dprior, 4. corr.DNprior and 5. dist.DNprior (proposed). Reconstruction volume sizes $|V|$: teapot 1: $150mm \times 200mm \times 80mm$; teapot 2: $120mm \times 200mm \times 120mm$; doll: $125mm \times 125mm \times 70mm$; billiard: $60mm \times 60mm \times 35mm$; cup: $125mm \times 125mm \times 70mm$; vase: $190mm \times 130mm \times 120mm$; mannequin: $210mm \times 160mm \times 100mm$; teddy: $170mm \times 180mm \times 80mm$. Integration: NoInt shown in flat shading.
Figure 3.15: Reconstruction of real static datasets continued (frontal view, smooth shading). Other parameters as in Figure 3.14.
3.5. Evaluation

Figure 3.16: Reconstruction of real static datasets continued (normal maps). Other parameters as in Figure 3.14.
Figure 3.17: Reconstruction of real static datasets continued (side view and smooth shading). Other parameters as in Figure 3.14.

Figure 3.18: Pseudo-geometry elimination using patch-based averaging. Feature-based patch sizes (in pixels): teapot 1: 40 × 40; teapot 2: 40 × 40; billiard: 60 × 60; vase: 55 × 55.
3.5. Evaluation

Integration method

In standard ML HS the final surface is formed using FC to perform direct integration of the normal field without taking point depths into account. PR alternatively is a more holistic integration approach using depth and normal estimates jointly. Initialised by the vertex 3D location, the method fits the surface as the boundary between the interior and exterior of the object in a way best reflecting the estimated normal field. Since the algorithm requires a full 3D point cloud, in this work the estimated 2.5D cloud is extended by injecting missing points within the spatial boundaries of the view’s contour that also defines the normals of the injected points. Finally, in the third integration method NoInt, explicit surface integration is avoided altogether simply meshing vertices geometrically into facets based on proximity defined by belonging to 4-connected neighbourhoods.

Taking CtF Bayesian HS with dist.DNprior as the reconstruction method (parameters as in Section 3.5.2), Figure 3.19 compares the integration results by FC, PR and NoInt for all datasets (for completeness Appendix B. also contains the FC and PR integration results for all the other priors). Firstly, one can observe the global shape distortion introduced by FC (e.g. the side views of the spherical objects, the shear in the frontal view of teapot 2). The artefact is typical of FC due to its underlying principles of operation. The FC method works by enforcing integrability of the reconstructed gradient field by effectively finding its closest integrable representation in the frequency domain. This \textit{a posteriori} integrability check is different from enforcing integrability via the prior in vertex estimation. Besides being an integrable approximation only, a surface obtained by FC is susceptible to drift due to the gradual accumulation of numerical error in its direct normal field integration. So, FC is clearly not suitable as the integration method for the CtF Bayesian HS pipeline as it leads to deterioration of the point cloud quality.

With PR the global shape is a visibly undistorted reflection of the input point cloud thanks to the depth estimate support of photometric normal information. The PR surface however is not the truest representation of the point cloud as in fitting the surface PR alters the vertex position and orientation of the input. This leads to the loss of the estimated per-vertex photometric normals of the original point cloud, while those of the fitted surface appear not to facilitate the same level of feature resolution (e.g. slightly blurred dimples and eyebrows of the doll). Furthermore, with PR the input point cloud has to be made
watertight by injecting points on the occluded side which might lead to inflation of the integrated surface.

With NoInt the geometry of the surface is the truest representation of the point cloud without any post-processing preserving the exact vertex positions and orientations. In comparison to PR, NoInt has been shown to generate equally accurate, albeit less smooth, pure geometries best exposed in flat shading. However, in smooth shading, due to preservation of the true per-vertex photometric normals, the surfaces by NoInt have better defined features than those by PR (e.g. the doll’s features). Another advantage is that NoInt simplifies the pipeline considerably being a book-keeping exercise arranging vertices into facets rather than an optimisation procedure like PR and FC. The advantages of PR are a smoother pure geometry as was mentioned above as well as the ability to integrate lower resolution point clouds (NoInt requires high point cloud densities to avoid the jagged facet effect).

In conclusion, PR and NoInt are comparable in performance as surface integration methods. In the remainder of this work NoInt will be used for ease of processing as well as to preserve the best consistency of the surface to the input point cloud. Note that the proposed integration-free pipeline is facilitated by the coarse-to-fine framework boosting the point cloud densities and the tailored integrability prior de-emphasising post-reconstruction integration (poorer quality point clouds obtained with less suitable priors or without them in contrast will call for mitigation by explicit integration). PR on the other hand is kept as a valuable alternative to be used in case of low point cloud densities.
3.5. Evaluation

Figure 3.19: Comparative evaluation of three surface integration approaches (Frankot-Chellappa (FC), Poisson Surface Reconstruction (PR) and our integration-free (NoInt) meshing) on the real datasets. Reconstruction method: Bayesian HS with dist.DNprior (parameters as before).
3.6 Conclusion

3D geometry reconstruction has been a much researched topic in the last couple of decades. Yet, while showing impressive performance within their own scope of applicability, prior methods lack the freedom in the reflectance model of the reconstructed scene. To address the limitation, a 3D reconstruction framework was developed in this chapter of the thesis based on the previously little explored niche method of Helmholtz Stereopsis known for its independence of the reflectance model. Coarse-to-fine (CtF) Bayesian HS with integrability prior without explicit surface integration presented here is a fundamentally novel formulation of HS resulting in a drastically improved reconstruction accuracy and resolution on datasets complex both geometrically and photometrically also with a much reduced set of input images required.

The proposed framework consists of several novel contributions. Firstly, the standard ML formulation of HS is replaced by a novel MAP (Bayesian) formulation introducing neighbourhood support in surface point estimation. Secondly, to harness neighbourhood support a suitable prior is developed enforcing surface integrability via consistency between depth and normal information. Note that the prior is tailored to the unique ability of HS to generate both depths and normals. Thirdly, a coarse-to-fine framework for Bayesian HS is presented achieving previously unattainable point cloud densities and eliminating the need for explicit surface integration when the developed integrability prior is used. Hence the surface integration method is similarly tailored to the rest of the pipeline as only with the coarse-to-fine implementation for resolution and the integrability prior for accuracy of the point cloud can one avoid explicit integration.

The extensive evaluation on numerous real and synthetic datasets presented in this chapter validates several claims. To begin with, Bayesian formulation without a doubt outperforms ML HS. Further, the distance-based formulation of the proposed prior directly related to surface integrability has been universally shown superior to its correlation-based formulation and the classical one-sided depth and normal priors. With the least-restrictive regularisation basis of integrability defining the smoothness cost the results exhibit both global and local accuracy, high resolutions and robustness against intensity noise. In a comparative evaluation it has been shown that the integration-free approach avoids the adverse effects of global shape distortion and oversmoothing that explicit integration techniques are
3.6. Conclusion

prone to. Quantitatively, the sub-millimetre and (sub-)\(^1\) centimetre accuracy on noiseless and heavily corrupted synthetic data respectively are encouraging results.

CtF Bayesian HS with integrability prior without explicit integration is a universally applicable reconstruction core as the prior ultimately seeking surface continuity only is the least restrictive, subject to perhaps only to the limitation of not being able to deal with mathematical singularities such as the step and Dirac delta transition functions. As the reconstruction core inherits the acquisition method from the original Helmholtz Stereopsis, in this chapter the scope has been limited to static scenes only. Yet the property of much reduced cardinality of input reciprocal image set with Bayesian HS opens up avenues for introducing Helmholtz Stereopsis for dynamic scene reconstruction for the first time. The next chapter presents Colour Helmholtz Stereopsis (CL HS) - the first ever methodology for dynamic scene reconstruction using HS that features 1. wavelength-multiplexing for acquisition addressing the input capture latency shortcoming of standard HS and 2. the presented Bayesian HS for reconstruction with the reciprocal pair count reduced by the adopted acquisition procedure to the bare minimum of three. The seamless integration of the reconstruction core proposed in this chapter into the dynamic scene reconstruction approach of the next chapter is a testament of its modular nature with applicability beyond improving static reconstruction of the original HS pipeline.

\(^1\)depending on the scale of reconstruction
Chapter 4

Colour Helmholtz Stereopsis for Dynamic Scenes with Uniform Chromaticity
In Chapter 3 Coarse-to-fine (CtF) Bayesian Helmholtz Stereopsis with integrability prior was proposed and evaluated as a novel framework for high resolution accurate geometric reconstruction of scenes with complex *a priori* unknown reflectance properties. The method yielded previously unattainable accuracies on datasets of high photometric and geometric complexity, also at a much reduced cardinality of input image set compared to prior art Helmholtz Stereopsis (HS) (8 as opposed to 18 reciprocal pairs in [101]). However, these reciprocal pairs were captured in a sequential acquisition procedure by rotating the scene relative to the set-up on a turn-table in fixed increments to ensure interchanging of the camera and light source exactly eight times. This shortcoming of sequential reciprocal pair acquisition is not addressed in Chapter 3 limiting the applicability of the highly advantageous method to static scenes only.

In this chapter, a novel framework of Colour Helmholtz Stereopsis (CL HS) is proposed which for the first time generalises HS to dynamic scenes while, as its reconstruction core, incorporating CtF Bayesian Helmholtz Stereopsis (henceforth Coarse-to-fine Bayesian Helmholtz Stereopsis solved by MRF optimisation (CtF MRF HS)) with integrability prior. The key idea is to modify the acquisition of standard HS from sequential white light to simultaneous multi-spectral where three RGB cameras each receive signal from two light sources of different frequencies, while being collocated with the third light source emitting its own frequency. Due to the use of physically different multi-spectral light sources and cameras, there will be signal inconsistencies which are resolved in the proposed framework by means of a tailored photometric calibration procedure. Further, the framework is complete with the mutually tailored reconstruction and surface integration algorithms compatible with the input of a drastically reduced number of reciprocal pairs. In summary, the contribution of Chapter 4 is a novel HS-based method uniquely allowing accurate high-resolution reconstruction of photo-geometrically complex dynamic scenes with *a priori* unknown reflectance using only three images. More specifically, the novelty lies in 1. the novel multi-spectral acquisition set-up for instantaneous capture for HS; 2. a tailored photometric calibration procedure for signal consistency with the proposed set-up; and 3. integration of the highly suitable CtF MRF HS reconstruction core from Chapter 3 to ensure reconstruction accuracy with very few images.

In this chapter we present CL HS for surfaces with uniform chromaticity whereas the ex-
tensions to the framework permitting arbitrary spatially-varying chromaticity of the reconstructed surface are deferred to Chapter 5. The chapter begins with Section 4.1 containing the background theory of the state-of-the-art White Light Helmholtz Stereopsis (WL HS) for static scenes featuring the camera/light source calibration procedure from [106]. Section 4.2 subsequently develops the theory of the proposed CL HS generalising HS to dynamic scene reconstruction. To achieve that, every stage of the pipeline is revisited, from acquisition principles for calibration and reconstruction to the formulation of the reconstruction problem itself. The evaluation in Section 4.3 contains a comprehensive error analysis of the approximations made in the acquisition stage of CL HS as well as the results at the calibration and final stages of the pipeline. The performance of CL HS is evaluated both quantitatively and qualitatively on both static and dynamic complex scenes with uniform chromaticity matching any chosen reference. The performance on the targeted datasets, such as for instance the deformation of a blank laminated paper sheet, is remarkable with high fidelity geometries in the face of highly specular glare. Also in the evaluation, a validation is provided in favour of the choice for the CtF MRF HS reconstruction core as well as for the need for photometric calibration of multi-spectral equipment.

4.1 White Light Helmholtz Stereopsis (WL HS)

Traditional HS described and built upon in Chapter 3 relies on sequential acquisition with white light illumination. The acquisition set up conventionally consists of just one camera and one light source suspended over the scene. For reciprocal image pair acquisition the positions of the camera and the light source must be mutually interchanged exactly. The easiest way to achieve that, and the one commonly used in previously published work on White Light Helmholtz Stereopsis (WL HS), is to rotate the scene on a turn-table relative to the equipment. Such a procedure is of course equivalent to physical equipment relocation to satisfy reciprocity. In the remainder for simplicity of visualisation, let us adopt the conceptual model of the camera and the light source physically changing positions for reciprocal acquisition.

As illustrated in Figure 4.1, for intensity sampling in the acquisition procedure of standard WL HS a perspective camera $C$ and a light source $S$ are initially centred at $c_1$ and $c_2$ respectively when sampling surface point $x$. Subsequently, the two are swapped so that
Chapter 4. CL HS for dynamic scenes with uniform chromaticity

\( \mathcal{C} \) is at \( \mathbf{c_2} \) and \( \mathcal{S} \) is at \( \mathbf{c_1} \). So, points \( \mathbf{c_1} \) and \( \mathbf{c_2} \) host both a camera and a light source at some stage in the acquisition process of a single reciprocal pair. The notion of such a camera/light source pair at a single location is first formalised as the Helmholtz camera abstraction by Jankó et al. [106]. A Helmholtz camera \( \mathcal{R} \) is a hybrid scene sampling entity of a collocated camera and light source, which acts as a camera in one reciprocal image and then as a source of illumination in the other. Let us stress that in traditional sequential WL HS, the camera/light source collocation is achieved temporally by their mutual swap rather than physically.

Helmholtz camera \( \mathcal{R} \) is photometrically characterised as both a camera and a light source by respectively its sensitivity function \( \sigma \) and the radiance function \( \rho \). Both \( \rho \) and \( \sigma \) are spatially varying as a function of ray \( \mathbf{v} \) from the surface point \( \mathbf{x} \) to \( \mathcal{R} \) (Figure 4.1). The image formation equation of standard HS as described in Related Work relates the measured intensity to the sampling geometry, radiance and the BRDF of the surface point:

\[
i_1 = \rho_{iso} f_r(\mathbf{v_2}, \mathbf{v_1}) \frac{\mathbf{n} \cdot \mathbf{v_2}}{\|\mathbf{c_2 - x}\|^2}
\]

Recall that a spatially constant radiance distribution \( \rho_{iso} \) of an isotropic light source is assumed in the equation while sensor sensitivity is not modelled at all with the implicit assumption of the uniform unity distribution. Furthermore, \( \rho_{iso} \) is taken to be the same for both images in the reciprocal pair. These assumptions are a simplification of the reality of how intensities are affected by the photometric characteristics of the sampling set-up, specifically the potentially spatially-varying radiance and camera sensor sensitivity distri-

![Figure 4.1: Reciprocal intensity sampling in WL HS.](image)
4.1. White Light Helmholtz Stereopsis (WL HS)

butions different per Helmholtz camera. Let us define two Helmholtz cameras (or \((C, S)\) pairs), \(R_1 = (C_1, S_1)\) and \(R_2 = (C_2, S_2)\), located at \(c_1\) and \(c_2\) respectively as in Figure 4.1. In the formation of the first reciprocal image \(I_1\) of the pair, the sensitivity \(\sigma_1(v_1)\) of \(R_1\) in the role of \(C_1\) and the radiance \(\rho_2(v_2)\) of \(R_2\) acting as \(S_2\) are involved in the more physically accurate image formation equation from [106]:

\[
i_1 = \frac{\rho_2(v_2)\sigma_1(v_1) f_r(v_2, v_1) v_2 \cdot n}{\|c_2 - x\|^2}, \tag{4.2}
\]

Intensity \(i_2\), which is the projection of \(x\) in the second image of the reciprocal pair, is obtained by interchanging the vector indices 1 and 2 in Equation 4.2. As also shown by Jankó et al. in [106], the generalised formulation of the image formation equation logically results in the following updated formulation of the HS constraint:

\[
\left(\frac{\mu_1(v_1)i_1}{\|c_1 - x\|^2} v_1 - \frac{\mu_2(v_2)i_2}{\|c_2 - x\|^2} v_2\right) \cdot n = 0 \tag{4.3}
\]

where \(\mu\) is the spatially varying photometric calibration parameter of a Helmholtz camera:

\[
\mu_k(v_k) = \frac{\rho_k(v_k)}{\sigma_k(v_k)}, \quad k = 1, 2 \tag{4.4}
\]

With one spatial distribution of the parameter per each Helmholtz camera involved, any HS set-up can be considered fully photometrically calibrated, jointly in terms of both the spatially varying camera sensor sensitivities and light source radiances. The practical significance of parameter \(\mu\) transcends the label of per-ray effective sensitivity from [106] and hence, due to its descriptiveness of the photometric characteristics, it is rather viewed in this dissertation as the spatially varying photometric calibration parameter of the equipment. Section 4.1.1 discusses the process of photometric calibration developed by Jankó et al. in [106] for estimation of the spatial distribution of parameter \(\mu\) of Helmholtz camera \(R\) in sequential WL HS. The procedure can be summarised in four stages. Firstly, a set of expressions linking two sample points in the spatial distribution of \(\mu\) is obtained from pairs of reciprocal image samples sharing the same illumination ray. Secondly, the expressions are used to constrain a set number of control points of the spatial distribution using the bi-cubic interpolation kernel. Thirdly, all the constraining expressions are substituted into Equa-
tion 4.3 and, given a known surface normal, the resultant well-constrained linear system is solved for the control points using the least-squares algorithm. Finally, a continuous spatial distribution of $\mu$ is computed by bi-cubic interpolation between the estimated control points.

Jankó et al. also validate the effectiveness of the proposed calibration procedure measuring the rms deviation from the expected orthogonality between the normal to a localised plane and the constraint vector $w$ from Equation 4.3:

$$w = \left( \frac{\mu_1(v_1)_{i_1}}{\|c_1 - x\|^2} v_1 - \frac{\mu_2(v_2)_{i_2}}{\|c_2 - x\|^2} v_2 \right). \quad (4.5)$$

Through the proposed photometric calibration, the rms error has been shown to decrease from $\pm 9.1^\circ$ to just $\pm 0.33^\circ$. (The corresponding mean values reported were $-1.1^\circ$ and $-0.58^\circ$ before and after calibration respectively). The residual rms error is attributed by Jankó et al. to the limitations in the precision of their calibration set-up.

### 4.1.1 Photometric Helmholtz camera calibration in WL HS

**Procedure.** The procedure for photometric calibration in WL HS as proposed by Jankó et al. in [106] computes spatial distribution of the coefficient $\mu$ from Equation 4.4 by making use of purposely inserted into the scene calibration planes. The original procedure utilises a highly controlled set-up with a motorised turn-table. The equipment to be calibrated (a camera and a light source 60 cm apart) is suspended about 1 m above the turn-table. The calibration board simulating the calibration planes is positioned on the turn-table and can be moved in fixed vertical increments relative to the calibrated equipment. An exact 180° scene rotation on the turn-table around the vertical axis turns a single camera/light source pair into a pair of Helmholtz cameras ($R_1, R_2$) by virtually interchanging their positions. To guarantee Helmholtz reciprocity through rotation instead of the physical collocation of the camera and the light source, Jankó et al. stress the importance of centring the turn-table carefully so that the rotation axis bisects the horizontal line between the camera and the light source at 90°. The centring procedure is performed manually. Since the Helmholtz cameras in this case are virtual, i.e. obtained through the turn-table rotation, every calibration frame consists of two shots to capture the response of the Helmholtz
camera in both its camera and light source capacity.

The photometric procedure requires capturing calibration planes in different positions relative to the Helmholtz camera. The translation of the calibration board is performed in fixed vertical increments parallel to the carefully centred rotation axis of the turn-table (see the diagram in Figure 4.2 for the photometric calibration geometry of Jankó et al. for WL HS). In this configuration the normal to the calibration plane remains unchanged.

The calibration acquisition procedure is not easy from the practical point of view as it requires 1. precise vertical displacement of the calibration plane in space; 2. a mechanical system to achieve reciprocity and 3. careful manual centring of the turn-table’s rotation axis relative to the calibrated acquisition equipment. This highly controlled calibration set-up of Jankó et al. is suitable for calibration of a single \((R_1, R_2)\) pair but does not scale well to simultaneous calibration of multiple Helmholtz camera pairs in arbitrary positions as the calibration plane orientation is unlikely to be optimal for multiple pairs at the same time unless restrictions are imposed on where the calibrated Helmholtz cameras may be. For example, in full 3D reconstruction acquisition scenarios with the object surrounded by cameras, the calibration planes in a fixed (apart from vertical displacement) configuration can never be visible in all views.

![Figure 4.2: Geometry of the photometric calibration procedure by Jankó et al. for WL HS.](image)

**Mathematical details.** In the WL HS photometric calibration [106] of Helmholtz camera \(R_1\) another Helmholtz camera \(R_2\) is used to link HS constraints sampled on the calibration plane in its different positions. The constraints are linked via the ray of incident illumination \(v_2\) (Figure 4.2). For every position \(j\) of the calibration plane \(\Pi_j\) a ratio \(\kappa\) of parameters \(\mu_1\) and \(\mu_2\) is established. Parameters \(\mu_1\) and \(\mu_2\) correspond to the Helmholtz cameras \(R_1\) and
and are sampled at a surface point \( x \) where rays \( \mathbf{v}_1 \) and \( \mathbf{v}_2 \) intersect (see Figure 4.2).

For WL HS Jánko et al. obtain the following expression for \( \kappa \):

\[
\kappa(\mathbf{v}_1, \mathbf{v}_2 | \Pi_j) = \frac{\mu_1(\mathbf{v}_1)}{\mu_2(\mathbf{v}_2)} = \frac{\mathbf{n}^\top \mathbf{v}_2 \| \mathbf{c}_1 - \mathbf{x} \|^2_i}{\mathbf{n}^\top \mathbf{v}_1 \| \mathbf{c}_2 - \mathbf{x} \|^2_i} \quad (4.6)
\]

Subsequently, point \( x \) on plane \( \Pi_j \) is transferred onto the plane in the new position \( \Pi_{j+1} \) by finding the intersection \( x' \) of ray \( \mathbf{v}_2 \) with \( \Pi_{j+1} \). Hence for plane \( \Pi_{j+1} \) the ratio \( \kappa(\mathbf{v}_2^2, \mathbf{v}_2 | \Pi_{j+1}) = \frac{\mu_1(\mathbf{v}_2)}{\mu_2(\mathbf{v}_2)} \) is established sharing the denominator with the corresponding relationship of plane \( \Pi_j \). The shared denominator allows to obtain a relationship between \( \mu_1(\mathbf{v}_2^2) \) and \( \mu_1(\mathbf{v}_1) \) (two pixel locations \((u_1, v_1)\) and \((u_2, v_2)\) corresponding to rays \( \mathbf{v}_1 \) and \( \mathbf{v}_2^2 \) in the spatial photometric parameter distribution of \( \mathcal{R}_1 \)):

\[
r_1(\mathbf{v}_1, \mathbf{v}_2^2) = \frac{\kappa(\mathbf{v}_1, \mathbf{v}_2 | \Pi_j)}{\kappa(\mathbf{v}_1^2, \mathbf{v}_2 | \Pi_{j+1})} = \frac{\mu_1(\mathbf{v}_1)}{\mu_1(\mathbf{v}_1^2)}. \quad (4.7)
\]

A constraint in the form of Equation 4.7 establishes a relationship between pairs of points indexed \( k \) and \( k' \) and defined by vectors \( \mathbf{v}_k^1 \) and \( \mathbf{v}_{k'}^1 \) in the spatial distribution of \( \mu_1 \) characterising Helmholtz camera \( \mathcal{R}_1 \) photometrically:

\[
\mu_1(\mathbf{v}_k^1) = \mu_1(\mathbf{v}_{k'}^1) r_1(\mathbf{v}_k^1, \mathbf{v}_{k'}^1) \quad (4.8)
\]

As all the entities in Equation 4.8 are positive, it can be re-written as a sum by taking the logarithm:

\[
\lambda_1(\mathbf{v}_k^1) = \lambda_1(\mathbf{v}_{k'}^1) + \delta(\mathbf{v}_k^1, \mathbf{v}_{k'}^1) \quad (4.9)
\]

where \( \lambda \) and \( \delta \) correspond to the logarithm of \( \mu \) and \( r \) respectively. Entities \( \lambda_1(\mathbf{v}_k^1) \) and \( \lambda_1(\mathbf{v}_{k'}^1) \) are two samples in the logarithmic spatial photometric parameter distribution linked via Equation 4.9.

In order to estimate the continuous distribution, a regular spatial grid of \( N \) control points is introduced together with the assumption of spatial smoothness: i.e. discrete control points of the distribution are computed and interpolated between. The choice was made in [106] to model smoothness using the bi-cubic interpolation kernel. Any distribution sample \( \lambda(\mathbf{v}_k^1) \) can be expressed as a linear combination of \( N \) control point values based on the rules of bi-cubic interpolation: \( \lambda(\mathbf{v}_k^1) = a_1 \lambda^1 + a_2 \lambda^2 + \cdots + a_N \lambda^N \). Equally valid would be an
equivalent expression for the difference of two samples \( \lambda_1(\mathbf{v}_k^1) \) and \( \lambda_1(\mathbf{v}_k^1') \) derived from their individual linear combinations of control points. The additional knowledge of the relationship between the samples from Equation 4.9 precipitates in a single constraint on the control points:

\[
(a_k - a_{k'}) \begin{bmatrix} \lambda^1 & \lambda^2 & \ldots & \lambda^N \end{bmatrix}^\top = \delta(\mathbf{v}_1^k, \mathbf{v}_1^{k'}) \tag{4.10}
\]

where \( a_k \) and \( a_{k'} \) are the interpolation coefficient vectors of linked samples \( \lambda_1(\mathbf{v}_k^1) \) and \( \lambda_1(\mathbf{v}_k^1') \) respectively. By accumulation of an arbitrarily large number \( M \) of such constraints \( (M \gg N) \) the linear system of control point equations can be made over-constrained and hence easily solvable. Hence the motivation behind the estimation strategy of Jankó et al. involving distribution discretisation using control points with the spatial smoothness assumption is a problem simplification by variable reduction permitted through prior knowledge. The accuracy of control point estimation will depend on the number \( M \) of constraints used whereas the fidelity of the continuous distribution will depend on the density of the regular grid of control points as well as the complexity of the chosen regularisation kernel.

In the context of HS, the spatial photometric parameter distributions of the two Helmholtz cameras involved in the normal constraint in Equation 4.3 have to be consistent with each other. To achieve this consistency, Jankó et al. compute the control points of both distributions as part of the same homogeneous linear system with \( 2 \times N \) unknown control points, \( N \) for each camera in the pair. The solved linear system is derived from the normal constraint in Equation 4.3. A single constraint on the double set of linear control points \( \mu^{2N} \) based on Equation 4.3 is:

\[
\left( \frac{i_1}{\| \mathbf{c}_1 - \mathbf{x} \|^2} \mathbf{v}_1^\top \mathbf{n} \right) \alpha_1^{2N} - \left( \frac{i_2}{\| \mathbf{c}_2 - \mathbf{x} \|^2} \mathbf{v}_2^\top \mathbf{n} \right) \alpha_2^{2N} \cdot \mu^{2N} = 0 \tag{4.11}
\]

where 1. \( \mu^{2N} = \begin{bmatrix} \mu_1^1 & \mu_1^2 & \ldots & \mu_1^N & \mu_2^{N+1} & \ldots & \mu_2^{2N} \end{bmatrix}^\top \) and 2. \( \alpha_1^{2N} \cdot \mu^{2N} \) and \( \alpha_2^{2N} \cdot \mu^{2N} \) are two samples from spatial distributions \( \mu_1 \) and \( \mu_2 \) respectively determined by vectors \( \mathbf{v}_1 \) and \( \mathbf{v}_2 \) expressed as an interpolation from the control points. Note that the interpolation coefficient vector \( \alpha_1^{2N} \) has \( N \) trailing zeros in its \( 2N \) entries, while in \( \alpha_2^{2N} \) there are \( N \) leading zero entries, with the zeros in each case corresponding to the control points of the
other camera’s distribution. $M$ constraints as in Equation 4.11 can be accumulated into a homogeneous linear system:

$$A_{M \times 2N} \mu^{2N} = 0. \quad (4.12)$$

The resultant linear system is well-constrained, as one has direct control over the number of variables (i.e. control points) and constraining expressions, and is hence easily solvable by the least-squares optimisation algorithm. Note that in this formulation for joint photometric distribution estimation of $\mathcal{R}_1$ and $\mathcal{R}_2$ the linear values $\mu$ can be obtained directly, instead of the logarithmic values $\lambda$ derived previously. Having computed the control points for both distributions, the continuous solutions are obtained using bi-cubic interpolation.

**Overall assessment.** Let us point out a couple of aspects in the formulation of the discussed calibration algorithm by Jankó et al. that require further attention. Firstly, the algorithm has been validated for photometric calibration via just one other Helmholtz camera i.e. in a single Helmholtz camera pair. The choice to restrict to just one pair is understandable given the sequential nature of their acquisition set-up. However, as Jankó et al. point out, the problem in this case will become ill-posed for the lack of samples interlinking constraint chains in the direction orthogonal to the epipolar lines. The problem has the maximum effect when the epipolar lines of the two cameras are parallel (see Section 4.2.3 for a more detailed analysis and illustration of the phenomenon). In more favourable circumstances the strong bi-cubic regulariser employed in the algorithm seems to provide enough support for a consistent solution according to their results. As strong regularisation may have side-effects such as over-smoothing, the interpolation kernel choice is the second point that is worth re-consideration. Thirdly, while undoubtedly implicitly enforcing mutual consistency of spatial parameter distributions of $\mu_1$ and $\mu_2$ of respectively $\mathcal{R}_1$ and $\mathcal{R}_2$ in the joint control point estimation of Equation 4.11, the vital importance of the consistency without which no valid HS constraint in Equation 4.5 can be formulated is not explicitly discussed by Jankó et al.. Further, the impracticalities of the highly involved sequential acquisition procedure with specialist mechanical equipment, calibrated vertical plane displacements, manual centring etc. should be addressed particularly where the choices made limit scalability of the approach to simultaneous calibration of multiple Helmholtz cameras due to the discussed framing issues of the calibration plane and/or acquisition sequentiality. Similarly note that the camera placement with the acquisition
procedure is undesirably restricted to the circular path around the rotation axis of the turn-table. All these issues will be addressed later on in this chapter when formulating a new procedure for photometric calibration of a Helmholtz camera specifically developed for Colour Helmholtz Stereopsis (CL HS).

4.2 Colour Helmholtz Stereopsis (CL HS)

Colour Helmholtz Stereopsis (CL HS) is a novel approach and one of the key contributions of this thesis for the first time generalising Helmholtz Stereopsis (HS) to dynamic scene reconstruction. In this section, firstly, the acquisition principles of the proposed technique are introduced, which are fundamentally altered relative to White Light Helmholtz Stereopsis (WL HS). Secondly, CL HS is formalised in a novel formulation reflecting the employed multi-spectral acquisition. Thirdly, a generalisation of photometric calibration to multi-spectral Helmholtz cameras is presented. An in-depth analysis and the resultant improvements of the calibration procedure are also given. One important improvement pertains to a more practical capture of calibration data requiring only low cost off-the-shelf equipment (i.e. a hand-held uniformly coloured planar object) without the need for its accurate displacement. Furthermore, the computational complexity of the calibration data processing has been reduced (e.g. a lower order interpolation kernel, smaller linear systems etc.) - the change partially facilitated by engaging multiview support beyond a single pair of Helmholtz cameras.

![Figure 4.3: Reciprocal intensity sampling in CL HS.](image-url)
4.2.1 Acquisition principles and formalisation

In contrast to WL HS, the acquisition set-up of CL HS, as shown in Figure 4.3, relies on the physical collocation of camera $C$ and light source $S$, instead of their temporal displacement, to realise each Helmholtz camera $\mathcal{R}$. In CL HS, there are two such physical collocation-based Helmholtz cameras $\mathcal{R}_1$ and $\mathcal{R}_2$ in a pair whose light sources $S_1$ and $S_2$ are characterised by spectra of distinct frequency characteristics, hence the name Colour Helmholtz Stereopsis (CL HS). Provided the cameras in the set-up allow to read the responses to $S_1$ and $S_2$ independently of one another, such a set-up allows simultaneous acquisition of both images in the reciprocal pair when $S_1$ and $S_2$ are on at the same time.

In practice, the set-up requires RGB cameras and light sources with red, green or blue illumination which can be achieved by using colour filters. It is important that the spectra of light sources match the spectral sensitivity characteristics of the corresponding camera sensors as closely as possible to minimise signal corruption by cross-talk.

In WL HS the constancy of sampling spectrum ensures consistent measurements within a reciprocal pair and the flatness of the spectrum guarantees a response regardless of the surface point colour. As a consequence, spectral properties of sampling illumination may be omitted from the formulation of the BRDF in WL HS resulting in the previously stated expression based on sampling geometry only: $f_r(v_2, v_1)$. In CL HS, two different sub-spectra of the flat spectrum are used for sampling within a single Helmholtz camera pair e.g. one Helmholtz camera in the pair may sample the red spectrum emitted by its partner camera, whilst itself transmitting the green spectrum received by the same partner. For a consistent frequency-independent response with such a multi-spectral acquisition set-up, chromaticity of the surface point matters. A white point will reflect both red and green spectra in the example Helmholtz camera configuration equally because of its flat reflectance spectrum as a function of sampling illumination frequency $\omega$. One can say that the illuminating spectra will be unaltered upon reflection. A point of any other chromaticity will absorb some wavelengths of sampling illumination and the extent of spectral absorption will not be the same for the two spectra in the configuration. In this case, the illumination spectra are spectrally modified by reflection. A reduced spectrum being integrated by the camera sensor will result in a lower intensity measurement. Since the spectral modification by reflection is not the same between channels, the result is signal inconsistency. In contrast, in
WL HS, the signal is consistent as the spectral modification upon reflection of the constant sampling illumination spectrum at a given point is always the same.

The dependence on surface chromaticity is made explicit by factoring sampling illumination frequency $\omega$ into the BRDF expression at $x$: $f_r(v_2, v_1, \omega)$. Let us decompose BRDF $f_r$ into its directional component $f_d(v_2, v_1)$, dependent only on the viewing vector and the vector of incident illumination, respectively $v_1$ and $v_2$, and the component related to the surface point chromaticity $p(\omega)$. Chromaticity is in a way the inherent colour or a property that describes spectrally the relative reflection and absorption characteristics without taking the brightness of colour into account. In the Hue Saturation Brightness (HSB) colour model, chromaticity can be thought parametrised by hue (i.e. the wavelength with the strongest response) and saturation (i.e. the width of the spectral lobe around the hue peak or, in other words, the purity of the dominant hue). Since the absolute strength of the integrated spectral response (i.e. brightness) is not a parameter, there are families of colours characterised by the same chromaticity e.g. all greyscale values including white form a single family. For a more detailed discussion on the notion of chromaticity the reader is referred to Chapter 5.

In the acquisition configuration of CL HS chromaticity is discretised as a triplet of reflectance coefficients corresponding to the three (typically RGB) illumination spectra. One member of the triplet $p_{1,2}$ at a given point is defined as the reflectance coefficient due to the inherent colour of the point projected to a pixel of camera $C_1$ when lit by light source $S_2$. Incorporating chromaticity $p_{1,2}$, the image formation equation (4.2) of WL HS from Section 4.1 is re-written for CL HS as:

$$i_1 = \rho_2(v_2)\sigma_1(v_1)p_{1,2}f_d(v_2, v_1)\frac{v_2 \cdot n}{\|c_2 - x\|^2}$$ (4.13)

Further, imposing Helmholtz reciprocity of the directional component $f_d(v_2, v_1) = f_d(v_1, v_2)$, for CL HS the normal constraint from Equation 4.3 consequently becomes:

$$\left(\frac{\mu_1(v_1)i_1}{p_{1,2}\|c_1 - x\|^2}v_1 - \frac{\mu_2(v_2)i_2}{p_{2,1}\|c_2 - x\|^2}v_2\right) \cdot n = 0$$ (4.14)

accommodating chromaticity components $p_{1,2}$ and $p_{2,1}$ of surface point $x$.

In this chapter, reconstruction of scenes with arbitrary but uniform chromaticity is tar-
Chapter 4. CL HS for dynamic scenes with uniform chromaticity

gated. In this case, chromaticity can be calibrated as part of the photometric parameter distribution: coefficients $p_{1,2}$ and $p_{2,1}$ constant across the surface can be bundled together with the spatially varying $\mu_1$ and $\mu_2$ respectively, effectively scaling the distributions. The joint calibration in practice means that there is a reference chromaticity ($p_{1,2}^{ref}$ and $p_{2,1}^{ref}$) shared by the reconstructed surface and the target plane in the photometric calibration procedure. As the result of the photometric calibration with the target of reference chromaticity, one estimates the relative spatial distributions $\mu_1' = \frac{\mu_1}{p_{1,2}^{ref}}$ and $\mu_2' = \frac{\mu_2}{p_{2,1}^{ref}}$ inherently calibrated chromatically for the reference. No further chromaticity calibration is necessary for any surface sharing the reference chromaticity, which includes a whole family of colours with the same inter-channel relationship as the reference. The colours may even vary in the scene as long as the variation is within the same chromaticity family e.g. greyscale colours only. The normal constraint from Equation 4.14 can be thus re-written as:

$$\left(\frac{\mu_1'(v_1)i_1}{\|c_1-x\|^2}v_1 - \frac{\mu_2'(v_2)i_2}{\|c_2-x\|^2}v_2\right) \cdot n = 0 \quad (4.15)$$

where $\mu_1' = \frac{\mu_1}{p_{1,2}^{ref}}$ and $\mu_2' = \frac{\mu_2}{p_{2,1}^{ref}}$ are the relative photometric parameter distributions. The fundamental difference in the corresponding constraint of WL HS (Equation 4.3) is the eliminated chromaticity scaling factor of the photometric distributions since in WL HS $p_{1,2}^{ref} = p_{2,1}^{ref}$.

The scope of this chapter is limited to surfaces with uniform reference chromaticity for initial validation of CL HS. The extension to arbitrary spatially varying chromaticity is made in the next chapter. Note that apart from the requirement of chromatic uniformity, no further restrictions are imposed on the scene since the reference chromaticity used in the photometric calibration can be chosen to match the reconstruction target. Furthermore, the directional component of the BRDF can be arbitrary and spatially varying and the scene can be either static or dynamic.

Given the physically different cameras and the multi-spectral nature of light sources, the need for accurate yet practical photometric calibration in CL HS is more acute than in WL HS. Section 4.2.3 details the new practical photometric calibration procedure generalised from the single mono-spectral Helmholtz camera pair configuration with sequential acquisition of Jankó et al. [106] to cater to CL HS as a multi-Helmholtz camera multi-spectral system with simultaneous acquisition.
4.2. Colour Helmholtz Stereopsis (CL HS)

4.2.2 Overview of the pipeline

Figure 4.4 shows schematically how the key stages of CL HS for reconstruction with uniform chromaticity tie together. In the pipeline the same multi-spectral acquisition procedure provides input imagery for both reconstruction and photometric calibration. Photometric calibration computes relative photometric distributions \( \mu'_1 \) and \( \mu'_2 \) for each Helmholtz camera pair \((R_1, R_2)\) in the set-up calibrating jointly for both the radiometric characteristics of the acquisition equipment and the reference chromaticity. The estimated distributions together are fed to the Helmholtz-Stereopsis-based reconstruction module where they are used to formulate sets of constraints as in Equation 4.15 for the reconstructed surface with the reciprocal image pairs acquired separately from calibration. Reconstruction outputs point clouds of oriented vertices which are integrated into a continuous surface as the final output. A more in-depth illustration and discussion of the flow with implementation details as well as inputs and outputs for each stage is given in Section 4.2.4.

4.2.3 Photometric Helmholtz camera calibration in CL HS

Procedure. In the original paper, the calibration was performed in a highly controlled environment with the plane translated in known vertical increments with a single camera and a light source suspended overhead and manually centred over the turn-table with the plane. For photometric calibration in the CL HS pipeline a more freehand approach is intro-
duced to enable simultaneous calibration of multiple physical-collocation-based Helmholtz cameras and to relax the need to perform a pre-defined motion. The physical collocation allows simultaneous calibration of the photometric behaviour as both a camera and a light source. The proposed acquisition procedure for photometric calibration, in contrast to the motorised turn-table of Jankó et al. in [106], features a hand-held calibration board moved within the reconstruction volume in a much less constrained manner.

As in WL HS, the board in various orientations simulates the calibration planes for the photometric calibration procedure. As discussed in Section 4.2.1, the planes in CL HS need to have a spatially uniform chromaticity of choice (i.e. the reference chromaticity) matching that of the surface to be reconstructed. Once thus calibrated, the set-up is inherently tailored for reconstruction of surfaces with uniform reference chromaticity only, although the surface colour may vary freely within the chromaticity family. In this work’s implementation of the proposed photometric calibration procedure, the chosen chromaticity of the calibration planes is characterised by an equal inter-channel relationship and their colour is approximately white. The choice of chromaticity calibrates the set-up for any greyscale scene. In theory, the planes may have colour non-uniformity as long as it is within the same chromaticity family. In practice, however, colour non-uniformity is undesirable as a cause of sampling inconsistencies due to geometric mis-projections within a reciprocal pair. Finally, with white being the brightest member of the greyscale chromaticity family, it is preferable to have the plane approximate white rather than any other (greyscale or otherwise) colour to maximise the reflectance response, and with it the signal-to-noise ratio, regardless of the illumination strength and spectrum.

Further, the proposed procedure for photometric calibration relies on accurate localisation of calibration planes, which is not trivial when the plane is moved around in an uncontrolled manner within the volume. For localisation the board bears three or more easily (and preferably automatically) detectable black (for better contrast against white) markers. Assuming the cameras in the set-up are calibrated geometrically, in terms of both intrinsic and extrinsic parameters [115], detection of at least three localisation markers in the image domain allows one to determine both the position and orientation of the plane in 3D.

**Mathematical details.** CL HS employs a set-up consisting of multiple Helmholtz camera pairs. For clarity, let us formulate photometric calibration of a Helmholtz camera in a single
pair of Helmholtz cameras. At the end the straightforward step to combine several pairs involving the calibrated Helmholtz camera into a single system will be clear.

Figure 4.5 schematically represents the calibration constraint formation process proposed for CL HS in a single Helmholtz camera pair, which is equivalent to Figure 4.2 except that the calibration planes in this case are not necessarily parallel.

Figure 4.5: Geometry of the proposed practical photometric calibration procedure for CL HS.

Unlike WL HS formulated with the absolute $\mu_1$ and $\mu_2$, $\kappa(v_1, v_2 \mid \Pi_j)$ in CL HS is the ratio of samples from the relative photometric parameter distributions $\mu'_1$ and $\mu'_2$ corresponding to the Helmholtz cameras $\mathcal{R}_1$ and $\mathcal{R}_2$ sampled at a surface point $x$ where rays $v_1$ and $v_2$ intersect at the calibration plane $\Pi_j$. The ratio is derived from Equation (4.14) and via $\mu'_1$ and $\mu'_2$ incorporates reference chromaticity coefficients $p_{1,2}^{ref}$ of the calibration plane:

$$\kappa(v_1, v_2 \mid \Pi_j) = \frac{\mu'_1(v_1)}{\mu'_2(v_2)} = \frac{p_{2,1}^{ref} \mu_1(v_1)}{p_{1,2}^{ref} \mu_2(v_2)} = \frac{n^T v_2}{n^T v_1} \frac{\|c_1 - x\|^2}{\|c_2 - x\|^2} i_2$$  (4.16)

The corresponding expression for calibration constraint $r_1(v_1, v_2^2)$ as stated for WL HS in Equation 4.7 for CL HS becomes:

$$r_1(v_1, v_2^2) = \frac{\kappa(v_1, v_2 \mid \Pi_j)}{\kappa(v_2^2, v_2 \mid \Pi_{j+1})} = \frac{\mu'_1(v_1)}{\mu'_1(v_2^2)} = \frac{p_{2,1}^{ref} \mu_1(v_1)}{p_{1,2}^{ref} \mu_1(v_2^2)} = \frac{\mu_1(v_1)}{\mu_1(v_2^2)}$$  (4.17)

where $\kappa(v_2^2, v_2 \mid \Pi_{j+1}) = \frac{\mu'_1(v_2^2)}{\mu'_2(v_2)} = \frac{p_{2,1}^{ref} \mu_1(v_2^2)}{p_{1,2}^{ref} \mu_2(v_2)}$ is the ratio at point $x'$ - the intersection of the illumination ray $v_2$ with another calibration plane $\Pi_{j+1}$. The chromaticity coefficient $p_{2,1}^{ref}$
under illumination of \( R_1 \) viewed by \( R_2 \) is a constant, equal at the locations sampled by \( v_1 \) and \( v_2^2 \), due to the given chromatic uniformity of the calibration plane. As the result, the calibration constraint of CL HS reduces to that of WL HS in Equation 4.7 because the absolute and the relative photometric distributions are identical except for the spatially constant scale factor of reference chromaticity. Given the result, Equations 4.8 and 4.9 that in WL HS describe the relationship between two absolute photometric distribution samples of \( \mu_1 \) defined by rays \( v_1^k \) and \( v_1^{k'} \), are valid for two relative distribution samples of \( \mu_1' \) as well. Note that \( \mu_1 \) and \( \mu_1' \) are estimated as identical by the procedure below unable to resolve the differentiating scale factor. However, the scale factor is explicit in Equation 4.16 where the ratio of \( \mu_1' \) to \( \mu_2' \) differs from that of \( \mu_1 \) to \( \mu_2 \) by the factor of \( \frac{p_{2,1}}{p_{1,2}} \). The scale factor allowing to differentiate between \( \mu_1 \) and \( \mu_1' \) is re-introduced in the parameter transfer procedure between pairs of photometric distributions in a reciprocal pair (see below) making use of Equation 4.16. The transfer results in a pair of properly scaled relative distributions

\[
\mu_1' = \frac{\mu_2}{p_{1,2}} \quad \text{and} \quad \mu_2' = \frac{\mu_2}{p_{2,1}}
\]

in the reconstruction constraint of Equation 4.15. In order to avoid complicating the notation and preserve consistency with Equations 4.8 and 4.9, in the remainder of the chapter relative photometric distributions will be implied throughout in the context of CL HS and denoted by \( \mu \) omitting the apostrophe.

Let us repeat the logarithmic form of the relationship from Equation 4.9 here for convenience, re-writing it as a difference:

\[
\lambda_1(v_1^k) - \lambda_1(v_1^{k'}) = \delta(v_1^k, v_1^{k'})
\]

(4.18)

where \( \delta(v_1^k, v_1^{k'}) = \ln(r_1(v_1^k, v_1^{k'})), \lambda_1(v_1^k) = \ln(\mu_1(v_1^k)) \) and \( \lambda_1(v_1^{k'}) = \ln(\mu_1(v_1^{k'})) \). Any two samples of the spatial photometric coefficient distribution \( \lambda_1(v_1^k) \) and \( \lambda_1(v_1^{k'}) \) can be expressed as a linear combination of members from the distribution’s control point set \( \lambda_1 = [\ln(\mu_1(v_1)), \ln(\mu_1(v_2)), \ldots, \ln(\mu_1(v_N))]^T \) according to the chosen interpolation kernel:

\[
\lambda_1(v_1^k) = a_i \lambda_1 \quad \text{and} \quad \lambda_1(v_1^{k'}) = b_i \lambda_1.
\]

In order to avoid over-regularisation, the weaker bilinear kernel is used in the proposed procedure instead of the bi-cubic kernel employed in the prior art for WL HS. From a set of \( M \) constraints in the form \( (a_i - b_i)\lambda_1 = \delta_i \) the following linear system of equations is formulated for \( R_1 \) in the Helmholtz camera pair with
4.2. Colour Helmholtz Stereopsis (CL HS)

N control points of the sought distribution as unknowns:

\[
(A - B)^\top (A - B) \lambda_1 = (A - B)^\top \Delta
\]  

(4.19)

where \( A = [a_0, a_1, \ldots, a_M]^\top \), \( B = [b_0, b_1, \ldots, b_M]^\top \) and \( \Delta = [\delta_0, \delta_1, \ldots, \delta_M]^\top \). Note that the complexity of the system in Equation 4.19 is only dependent on the number of unknowns \( N \), not the number of constraints \( M \). Permitted by this independence, at this stage in the photometric calibration of CL HS, the linear system in Equation 4.19 is expanded to incorporate additional constraint sets computed from the other Helmholtz pairs involving the calibrated \( \mathcal{R}_1 \) that are available in the configuration. The reasons for calibration using multiple Helmholtz camera pairs will be discussed in more detail later on in this section.

Having estimated the vector of logarithmic control points \( \lambda_1 \), the spatially varying distribution \( \mu_1 \) can be computed by applying bi-linear interpolation between the corresponding linear control points: \( m_1 = [\mu_1(v_1), \mu_1(v_2), \ldots, \mu_1(v_N)] \). If necessary, simpler forms of interpolation (vertical or horizontal only) can be used to extend the range of calibrated pixels.

**Parameter transfer for calibration consistency in a Helmholtz camera pair.** In the photometric calibration procedure for WL HS both cameras of the pair are solved for simultaneously, in a single linear system built from the constraints as in Equation 4.3. For reasons of scalability and potentially a finer spatial sampling discussed later on, the choice has been made in the proposed procedure for CL HS to calibrate each Helmholtz camera individually instead. In this case, to ensure mutual consistency in each Helmholtz pair the calibrated distribution of one camera must be transferred onto its partner camera. The absolute necessity of such mutual consistency was never stressed enough by Jankó et al. in [106] since in their case such consistency was implicitly automatically guaranteed through the global joint optimisation and hence never an issue.

In CL HS, the photometric calibration transfer in a Helmholtz camera pair is formulated as a linear system of a set of constraints based on Equation 4.16. Having computed the continuous distribution of \( \mathcal{R}_1 \) directly, the goal is to compute the transferred distribution \( \mu_{1,2} \) for \( \mathcal{R}_2 \) in a Helmholtz camera pair \( (\mathcal{R}_1, \mathcal{R}_2) \) which is scale consistent with its partner’s distribution \( \mu_1 \). The transferred distribution \( \mu_{1,2} \) is characterised by the set of control points \( m_{1,2} = [\mu_{1,2}(v_1^1), \mu_{1,2}(v_2^2), \ldots, \mu_{1,2}(v_N^2)] \). The transfer of a single sample from \( \mathcal{R}_1 \) to
Chapter 4. CL HS for dynamic scenes with uniform chromaticity

$\mathcal{R}_2$ can be expressed as:

$$\kappa(v_1, v^*_1 | \Pi_j)\mu_{1,2}(v^*_1) = \mu_1(v_1) \quad (4.20)$$

where $\kappa(v_1, v^*_1 | \Pi_j)$ is a single constraint for parameter transfer from $\mathcal{R}_1$ to $\mathcal{R}_2$. The expression for the constraint derived from the sampling geometry and observed intensities at the point defined by the intersection of vectors $v_1$ and $v^*_1$ on a given calibration $\Pi_j$ is:

$$\kappa(v_1, v^*_1 | \Pi_j) = \frac{n^T v^*_1 \|c_1 - x\|^2_i}{n^T v_1 \|c_2 - x\|^2_i} \quad (4.21)$$

Note that $v_1$ does not have to sample one of the control points of the spatial photometric parameter distribution $\mu_1$: any interpolated value of the distribution is a valid sample for parameter transfer onto $\mathcal{R}_2$. Just as in the direct calibration process, to maximise support in the transfer linear system solved, multiple calibration planes $\Pi_j$ are used in the transfer. Equally, non-control points $\mu_{1,2}(v^*) \notin m_{1,2}$ of the distribution $\mu_{1,2}$ also give rise to constraints as in Equation 4.21. For every transfer sample, the computed $\kappa$ constrains up to four control points from $m_{1,2}$ that are related to the sample point through the bi-linear interpolation kernel in the way defined in the N-dimensional constraint vector $k_i$. The transfer linear system solved is:

$$K^T K m_{1,2} = K^T M_1 \quad (4.22)$$

where $K = [k_1, k_2, ..., k_M]^T$ and $M_1 = [\mu_1^1, \mu_1^2, ..., \mu_1^M]$ is a set of corresponding known samples from the previously directly calibrated distribution $\mu_1$. The transfer procedure ensures mutual consistency of $m_{1,2}$ and the directly calibrated $m_1$. As such, values $m_{1,2}$ obtained through the transfer procedure described are different from $m_2$ that could have been computed by direct calibration of $\mathcal{R}_2$. Just as with the directly calibrated camera, the transferred control points are bi-linearly interpolated between to produce a continuous spatial distribution.

**Stable calibration in a multi-Helmholtz-camera-pair configuration.** CL HS has the inherent ability for simultaneous acquisition of three reciprocal pairs of images, involving three pairs of Helmholtz cameras. There is an advantage in calibrating a Helmholtz camera in multiple pairs, briefly mentioned by Jankó et al. in their work on calibration for WL HS,
4.2. Colour Helmholtz Stereopsis (CL HS)

Figure 4.6: Photometric calibration of a Helmholtz camera in multiple pairs: interlinking propagated constraint chains converging at the epipole.

which pertains to calibration robustness and stability. The advantage can be for the first time conveniently exploited in the context of CL HS.

When calibrating in a single pair, the calibration problem can easily become ill-posed as the relative positions of the two Helmholtz cameras constrain the sample set of the spatial photometric parameter distribution that can be described. As shown in Figure 4.6, when Helmholtz camera $\mathcal{R}_2$ with the centre at $c_2$ is calibrated via $\mathcal{R}_1$ with the centre at $c_1$, the set of calibration plane positions as shown will result in propagating chains of constraints converging in the epipole $e_{1,2}$. Jankó et al. point out that in the case of parallel propagating chains (i.e. parallel epipolar lines) the problem of spatial photometric calibration becomes under-constrained and the parameters can only be computed piecewise, along the parallel lines. The occurrence of such parallel epipolar lines is common, true for any pair of cameras arranged side-by-side framing the same scene as in that arrangement the epipole is located at infinity. In fact, even the constraint chains only tending to parallel, not intersecting within the scope of the image, already cause prohibitive instability of the spatial photometric calibration problem. The tendency to parallel constraint chains is present in any single pair configuration.

A multi-Helmholtz-camera-pair set-up can provide a solution for the problem. As shown in Figure 4.6, the issue can be easily resolved by introducing an extra Helmholtz camera $\mathcal{R}_3$ centred at $c_3$ provided its centre is not aligned with either of the other two cameras. $\mathcal{R}_3$ produces a new set of constraints stabilising the problem as the corresponding constraint chains are at a substantial angle (potentially tending to orthogonality) to the first set.
In WL HS obtaining suitable constraint chains from multiple pairs is not as straightforward as in CL HS due to the sequential acquisition with monochromatic illumination. Also in the procedure of Jankó et al. the use of strictly parallel calibration planes restricts the acquisition scope to a very particular Helmholtz camera pair configuration and would not be applicable in larger full 360° set-ups. Although theoretically not limited to a single pair, in practice Jankó et al. validate calibration using just two Helmholtz cameras. The instability of the problem in this case is resolved by means of a more complex bi-cubic regularisation. The approach works but inevitably makes assumptions about the photometric parameter behaviour in the direction orthogonal to the sampled chains which may not hold: the behaviour is not verified by sampling with another supporting camera. Note that in contrast the procedure for CL HS due to the extra Helmholtz camera support a weaker bi-linear kernel can be used which guards against over-regularisation.

In WL HS, the calibration is consistently performed on the greyscale sensor of the camera and the white spectrum of the light source. The sampled photometric distribution corresponding to the single virtually collocated camera/light source pair stays the same within any Helmholtz camera pair which in theory permits calibration in multiple pairs without any obstacle. In CL HS, for any physically-collocated camera/light source pair characterised by a single emitted wavelength, there are three photometric parameter distributions as each physical camera has an RGB sensor, i.e. three sensitivity distributions. Hence each such collocated pair is a multi-spectral Helmholtz camera. Only two of those sensitivity distributions are relevant per camera as the camera’s own wavelength will never be sampled by itself in CL HS. Hence, each multi-spectral Helmholtz camera in CL HS in theory gives rise to two Helmholtz cameras to be calibrated. If the channel sensitivity distributions of an RGB camera were required to be distinct, photometric calibration in CL HS would be limited to a single Helmholtz camera pair configuration as only one other Helmholtz camera in the set-up emits the wavelength of the channel defining the sensitivity of the calibrated Helmholtz camera while also sampling its illumination spectrum. However, the reasonable expectation of per camera colour independence in sensor manufacturing suggests the assumption that the per-channel sensitivity distributions of a given camera would be identical. Such an assumption allows to claim consistency of constraints for a given camera formed with illumination of different wavelengths in the set-up, thus facilitating calibration
using multiple Helmholtz camera pairs. The photometric calibration problem of CL HS is essentially reduced from six unknown photometric parameter distributions corresponding to six Helmholtz cameras (i.e. two sensitivity distributions and one radiance distribution per multi-spectral Helmholtz camera in a set-up of three) to just three (i.e. one radiance and one sensitivity distribution per multi-spectral Helmholtz camera). A similar channel equality assumption cannot be made about the radiance distributions in the set-up because of their dependence on the light source orientation in addition to the inherent parameters such as the illumination strength.

**Photometric calibration problem formulation in WL HS and CL HS.** In the original photometric calibration procedure for WL HS Jankó et al. advocate calibrating the entire configuration of two or more Helmholtz cameras in a single linear system. The advantage of joint calibration is the inherent mutual consistency of all estimates that are also individually more accurate due to a larger set of supporting constraints. However, the drawback of the approach is that the size of the system that needs to be solved grows with the number of Helmholtz cameras in the set-up.

In CL HS the full photometric complexity of configuration involves calibration of six, rather than three, Helmholtz cameras with six distinct sensitivity and three radiance distributions. In this case, global optimisation of Jankó et al. is not meaningful: since each distribution cannot be constrained involving more than two Helmholtz cameras due to the chromatic compatibility issue, global optimisation will only guarantee pairwise mutual consistency. With the assumption of per-camera equality of channel sensitivities, the photometric calibration problem in CL HS reduces to three distributions to be calibrated, which is equivalent to a three-Helmholtz-camera set-up in WL HS and could potentially be solved in a single linear system at the cost of computational complexity. To avoid this complexity, in this work an alternative approach described previously is used involving individual calibration of each multi-spectral Helmholtz camera by its chromatically distinct radiance and a generic (non-colour-dependent) sensitivity distribution. Then to ensure the essential mutual photometric calibration consistency within each reciprocal pair, the parameters are transferred to each partner Helmholtz camera. In other words, the alternative approach to global optimisation by Jankó et al. establishes mutual consistency of parameters explicitly *a posteriori* instead of inherently by joint estimation. The main advantages of the pro-
posed formulation of the photometric calibration problem are 1. its scalability, as unlike joint calibration, there is no limit on the number of Helmholtz cameras in the configuration due to their separate calibration and 2. potentially a higher resolution spatial parameter distribution since, calibrating one camera at a time, the distribution can be sampled more finely.

**Constraint acquisition for photometric calibration of CL HS.** As described previously, the procedure for constraint acquisition in the proposed photometric calibration procedure for CL HS varies considerably from the highly controlled set-up with vertical plane translation originally proposed by Jankó *et al.* In a multi-Helmholtz-camera-pair set-up of CL HS a greater range of plane motion patterns must be used to ensure mutual involvement of all cameras via constraints. In this section, some practical recommendations for the free-hand plane placement in the reconstruction volume will be discussed. Correct plane displacement patterns are also important to maximise the extent of spatial calibration coverage.

In order to illustrate the key concerns, let us consider constraint acquisition in a single pair with the calibrated Helmholtz camera $\mathcal{R}_1$ and the supporting Helmholtz camera $\mathcal{R}_2$ (Figure 4.7a.). The calibration volume of $\mathcal{R}_1$ is defined by the overlap of the visual cones of $\mathcal{C}_1$ and $\mathcal{C}_2$ (outlined in red in Figure 4.7). No point outside on the volume can ever be used for photometric calibration. The acquisition procedure is based on computing sets of related HS constraints on the calibration planes purposely inserted into the calibration volume. The HS constraints are related via the ray to the supporting camera (see Figure 4.5). Ideally, it is desirable to maximise the intersection of the inserted calibration planes and the calibration volume since as is illustrated in Figure 4.7a planes passing through the widest part of the volume give rise to more related constraint pairs and consequently a better spatial coverage of the photometric calibration of $\mathcal{R}_1$. For example, ray $r_1$ to calibration plane $\pi_2$ cannot form a pair of related HS constraints with plane $\pi_1$ because the intersection point of the propagation ray to the supporting camera $\mathcal{R}_2$ and the latter plane falls outside of the defined calibration volume, specifically outside of the visual cone of the calibrated $\mathcal{R}_1$. The same propagation ray however successfully links the HS constraints on planes $\pi_2$ and $\pi_3$ which are positioned deeper within the calibration volume. Equivalently, ray $r_3$ from $\mathcal{R}_1$ to the calibration plane $\pi_4$ is not a valid calibration sample as it intersects the plane outside of
4.2. Colour Helmholtz Stereopsis (CL HS)

Figure 4.7: Principles of practical photometric calibration for CL HS: a. calibration data scope and spatial coverage; b. recommended calibration data acquisition practices to maximise spatial coverage of calibration.

In practice, the spatial dimensions and orientation of the calibration volume are difficult to determine real-time during capture when the calibration is carried out. In addition, the calibration footage is captured for several pairs of Helmholtz cameras simultaneously, each with a different optimal calibration volume. Therefore, in the absence of the exact knowledge of the calibration volume, a strategy that has been found to work well in practice is to sweep a realistically large range of positions in depth. The sweep ensures that the widest dimension of the calibration volume is engaged at some point in the procedure at least and one is not calibrating at the extremities of the calibration volume.

Further, there is no need to maintain a fixed calibration plane orientation horizontally as in Jankó et al. As shown in Figure 4.7a, planes \( \pi_2 \) and \( \pi_3 \) of very different orientations both produce a valid pair of related HS constraints. In fact, given simultaneous calibration of several Helmholtz camera pairs the use of variable plane orientations is desirable as well as sweeping the plane in depth along the vector defined by its orientation (Figure 4.7b.) to maximise calibration on the outskirts of the calibration volumes. As shown in Figure 4.7a, planes \( \pi_2 \) and \( \pi_3 \), in the area of their largest divergence, often fail to form a constraint within the calibration volume for rays towards the outskirts of the visual cone of \( \mathcal{R}_1 \).
Figure 4.8: Acquisition set-up: a. the real acquisition set-up used b. schematic representation consisting of three multi-spectral Helmholtz cameras with the relevant parameters indicated.

The chance of forming a constraint for these points is much improved in Figure 4.7b, where a sequence of parallel calibration plane positions (e.g. $\pi_{1,a}, \pi_{1,b}, \pi_{1,c} ...$) displaced in depth is offered for a range of plane orientations (e.g. $\pi_1$). Note that depending on the ray position within the visual cone of $\mathcal{R}_1$ different plane orientations ($\pi_1, \pi_2, \pi_3 ...$) are required.

### 4.2.4 Implementation

In this section the realisation of the proposed CL HS in practice is discussed particularly focusing on the data acquisition set-up and the reconstruction framework in the core of the pipeline and drawing the reader’s attention to the characteristic properties of the tailored implementation.

**Acquisition set-up.** The acquisition set-up consisting of three RGB cameras $C_i$ and three light sources $S_i$ at positions $i = \{1, 2, 3\}$ devised to realise CL HS is shown in Figure 4.8. The cameras used are the professional Grass Valley Viper model. Sources $S_i$ are given different frequency characteristics by using dichroic colour filters\(^1\) - red, green and blue for maximal spectral separation. The filters are chosen to match the RGB spectral characteristics of the cameras as closely as possible. The schematic representation of the set-up in Figure 4.8b labels the multi-spectral Helmholtz cameras and relevant dimensions of the set-up for future reference. The collocation of cameras and light sources is approximated by placing them close next to each other. Note that for consistency mis-collocation is in

\[^{1}\text{manufacturer: Edmund Optics; size: 50\,mm square}\]
the same direction for all three pairs. An in-depth error analysis on the collocation approximation is presented in Section 4.3.1. Each collocated camera/light source pair (i.e. multi-spectral Helmholtz camera) partakes in two Helmholtz cameras resulting in the following specific configuration of six Helmholtz cameras: $C = \{R_{1,r}, R_{1,g}, R_{2,b}, R_{2,g}, R_{3,r}, R_{3,b}\}$

where each Helmholtz camera is defined as a light spectrum transmitter at position $\{1, 2, 3\}$ and a sensor of channel $\{r, g, b\}$. With the set-up three reciprocal image pairs per frame are simultaneously acquired, each pair uniquely characterised by a sub-set of two Helmholtz cameras from $C$. The Helmholtz cameras from $C$ are photometrically calibrated as described in Section 4.2.3 with the assumption of equal spatial RGB channel sensitivities per camera i.e. $\sigma_{1,r} = \sigma_{1,g}$ and $\sigma_{2,b} = \sigma_{2,g}$ and $\sigma_{3,r} = \sigma_{3,b}$. With this assumption the triplet of spatially varying photometric parameters $(\mu_1, \mu_2, \mu_3)$ fully photometrically describes the configuration. However, for pairwise parameter consistency within each reciprocal pair, the distribution of one Helmholtz camera is transferred onto the other forming the pair as described in Section 4.2.3, which results in a transferred distribution set, e.g. $(\mu_1, 2, \mu_2, 3, \mu_3, 1)$, for the three-reciprocal-pair system of CL HS.

Reconstruction framework and scope. The acquisition set-up generates three reciprocal pairs of images from which constraints in the form as in Equation 4.15 can be formulated. The unique choice of each light source spectrum determines the active channels of each camera and unambiguously defines the set of reciprocal images. In this chapter the scope is restricted to surfaces with uniform chromaticity matching the reference chosen at the photometric calibration stage: choosing $p_{1,2} = p_{1,2}^{ref}$ and $p_{2,1} = p_{2,1}^{ref}$ in Equation 4.14 defines the relative photometric parameter distributions in Equation 4.15 tailored to the reconstruction scope. Signal inconsistency introduced by spatially varying and/or non-reference surface chromaticities is not considered. CL HS generalisation to such unconstrained surfaces is deferred to Chapter 5.

The constraints as defined in Equation 4.15 are applicable in the original ML HS reconstruction framework as proposed by Zickler et al. [101]. However, ML HS is not a suitable option for the reconstruction stage of the CL HS pipeline because of the drastically reduced, specifically minimal, reciprocal pair set characteristic of CL HS. In Chapter 3, the shortcomings of ML HS with an 8-reciprocal pair input set (as opposed to 18 pairs used by Zickler et al.) were demonstrated exposing the proneness of the formulation to noise
due to the lack of regional support in depth assignment. Also in Chapter 3, Coarse-to-fine Bayesian Helmholtz Stereopsis solved by MRF optimisation (CtF MRF HS) with the optimal distance-based depth-normal consistency prior (dist.DNprior) enforcing integrability was developed and validated as a much more accurate and robust formulation of Helmholtz Stereopsis. In this chapter, in order to deal with the reconstruction challenge given just three reciprocal pairs available in CL HS, the previously proposed CtF MRF HS with dist.DNprior is incorporated into the pipeline. This optimality of prior through a holistic approach to reconstruction makes CtF MRF HS with dist.DNprior a reconstruction algorithm option perfectly tailored to the strict requirements of CL HS.

Further, as in Chapter 3, the sampling of the 3D space is done with an orthographic virtual camera and the output of the reconstruction process is a point cloud of oriented vertices. Intensity and photometric coefficient sampling in the framework is voxelwise rather than pointwise meaning that the measurement is averaged over a window containing projections of the eight corners of a voxel. Just before the sampling window starts converging to a single pixel (typically at the spatial resolution of $0.25 - 0.4 \text{ mm}$), the spatial resolution subdivision is stopped as the input image resolution limit is reached and increasing the reconstruction resolution spatially any further becomes pointless. This spatial resolution limit will be referred to as the spatial subdivision termination threshold.

In dynamic scene reconstruction each frame is reconstructed separately independently of the others. The settings for the initial resolution and the number of iterations in the coarse-to-fine framework would determine the level of final point cloud resolution and with it the most suitable surface integration back-end (see Figure 4.9). Note that the framework presented in this chapter targets reconstruction of dynamic scenes with a uniform chromaticity of choice (the reference) which are otherwise unconstrained in terms of the reflectance model.

**Overall pipeline for CL HS.** The pipeline in Figure 4.9 presents a detailed overview of CL HS for complex dynamic scenes with uniform chromaticity filling in crucial details into the schematic representation of Figure 4.4. The implementation choices made for reconstruction and integration stages are tailored to each other and to the minimalist input provided by the acquisition stage to maximise the quality of the output.

The input for calibration and reconstruction is obtained in the *acquisition stage* with the
Figure 4.9: A detailed CL HS pipeline for reconstruction of complex dynamic scenes with uniform chromaticity.
The photometric calibration stage produces a set of six photometric maps - three from direct calibration of multi-spectral \((R_1, R_2, R_3)\) with the assumption of \(\sigma_{1,r} = \sigma_{1,g}\) and \(\sigma_{2,b} = \sigma_{2,g}\) and \(\sigma_{3,r} = \sigma_{3,b}\) and the other three obtained by transfer from the directly calibrated distributions to ensure pairwise scale consistency within the reciprocal pairs. The reconstruction core stage of the pipeline has a feedback loop being a coarse-to-fine framework but can naturally be run with a single iteration. HS constraints are computed from the reciprocal pair intensity images taking the relative photometric parameter distributions into account as in Equation 4.15 (Section 4.2). Singular Value Decomposition (SVD) is performed on the constraints characterising every local depth hypothesis by a normal estimate with a measure of its quality. The MRF framework implements Bayesian HS with distance-based depth-normal consistency prior (dist.DNprior) from Chapter 3. The output of the reconstruction stage is a point cloud whose density depends on the final resolution of the coarse-to-fine reconstruction. If the resolution is low, the subsequent integration stage of the pipeline has to make use of an explicit integration algorithm, the most suitable of which is Poisson Surface Reconstruction (PR) utilising both depth and normal estimates. If the resolution is high, meshing can be achieved without explicit integration by ordering vertices into facets based on known geometric relationships within the reconstruction volume (NoInt) from Chapter 3. The final result is a continuous surface mesh.

### 4.3 Evaluation

The evaluation validates CL HS for complex dynamic scenes with uniform chromaticity through a series of quantitative and qualitative experiments on both real and synthetic data. For that purpose, the output at different stages of the pipeline is presented. The scenes used for validation in this section are all characterised by the greyscale chromaticity with an equal inter-channel relationship since a white plane was used for photometric calibration. The validation section is organised as follows. Firstly, the approximation made at the acquisition stage regarding the camera/light source collocation is thoroughly evaluated in Section 4.3.1 by looking at the consequences of the miscollocation on the accuracy of depth and normal estimates in a series of synthetic simulations. Secondly, the spatial photometric parameter distributions computed for the real acquisition set-up are presented and
discussed in Section 4.3.2. Thirdly, the accuracy of the pipeline is analysed quantitatively and qualitatively on real static datasets in Section 4.3.3. The evaluation section culminates in Section 4.3.4 with the ultimate validation of the proposed pipeline by reconstruction of dynamic scenes, featuring geometrically intricate non-rigid deformation of uniformly white photometrically complex objects.

4.3.1 Collocation error analysis

The collocation of camera $C$ and light source $S$ in the proposed acquisition set-up (Figure 4.8a) is not exact, with the pair placed side-by-side only approximating a Helmholtz camera. In this miscollocation the configuration deviates from the aimed for theoretical model of three multi-spectral Helmholtz cameras. Collocation of the optical camera centre and the centre of the light source is challenging even with specialist hardware meaning that the ability of the algorithm to tolerate some miscollocation error is a desirable property. As can be observed from the real data experiments in the following sections, the effect of miscollocation error on the CL HS results obtained with the CtF MRF HS reconstruction core is clearly not prohibitive and not even readily apparent. For a quantitative evaluation of the extent of collocation error, synthetic data with ground truth is therefore utilised in this section.

The simulated collocation error analysis consists of two parts:

1. normal error analysis: the study of the inherent normal estimate error that the reconstruction algorithm has to face due to miscollocation;

2. reconstruction with perturbed synthetic data: the reconstruction accuracy achieved by CtF MRF HS with dist.DNprior given the collocation error.

While part 1 allows to gauge the effect of miscollocation on individual normals and SVD residuals independently, part 2 illustrates the overall effect of the perturbation on both depth and normal estimates jointly optimised across the entire surface using the best performing reconstruction core. Together the experiments reveal both the immediate error due to miscollocation and the way the proposed reconstruction core manages the perturbed input. To begin with, the method for synthetic data generation is described.

**Synthetic data generation.** The synthetic set-up used for the collocation error simulation represented schematically in Figure 4.8b closely follows the real-life configuration with the
three (multi-spectral) Helmholtz cameras being in the same vertical plane in a roughly circular configuration with $r_{\text{cam}} = 1000\, \text{mm}$ and centred at the virtual camera, which is also the origin. The camera plane is at the distance $d_{\text{scene}} = 1200\, \text{mm}$ from the scene. Let us sample a single scene point $x$ in the 3D space with the surface normal $n$. The Helmholtz cameras $R_1$, $R_2$ and $R_3$ are located at angles $\angle R_1 = -90^\circ$, $\angle R_2 = 45^\circ$ and $\angle R_3 = 135^\circ$ to the horizontal (see Figure 4.8b.) and are centred at $c_1$, $c_2$ and $c_3$ respectively. In a reciprocal pair of images, the expected synthetic intensity measurements for the point are computed with the image formation equations such as the following for the camera pair $(R_1, R_2)$:

\[
\begin{align*}
i_1 &= I_{\text{max}} L_{\text{max}} f_r(v_2, v_1) \frac{v_2 \cdot n}{\|c_2 - x\|^2} \\
i_2 &= I_{\text{max}} L_{\text{max}} f_r(v_1, v_2) \frac{v_1 \cdot n}{\|c_1 - x\|^2}
\end{align*}
\]  

(4.23)

where $I_{\text{max}}$ is the maximum intensity level corresponding to the depth of the generated input image and $L_{\text{max}}$ is the maximum light intensity at the surface point ($L_{\text{max}} \leq 1$).

Let us in this analysis assume a 16-bit image with $I_{\text{max}} = 2^{16} - 1$ intensity levels and the maximum light intensity of 80% to avoid saturations: $L_{\text{max}} = 0.8$.

The choice of the BRDF function $f_r$ at $x$ is of importance because its nature will partially determine the system’s reaction to collocation perturbation. The commonly used Phong model [51] consists of the ambient ($f_a$), diffuse ($f_{\text{diff}}$) and specular ($f_{\text{spec}}$) components as the three constituent parts of reflectance:

\[f_{\text{phong}} = f_a + f_{\text{diff}} + f_{\text{spec}}\]  

(4.24)

Let us disregard the ambient component since Helmholtz Stereoscopy operates in the total absence of ambient illumination. Two distinct BRDF models are used for data generation: a purely Lambertian one with diffuse reflectance only and a specular model which is a realistic combination of diffuse and specular components.

The purely Lambertian BRDF is $f_r = f_{\text{diff}}$ where $f_{\text{diff}}$ is a constant and hence independent of the sampling geometry. The specular BRDF is defined as a weighted linear combination of a diffuse and a specular component: $\alpha f_{\text{diff}} + (1 - \alpha) f_{\text{spec}}$. If the point’s BRDF is specular, the surface reflectance will be strongly directionally dependent. The perfect mirror-like
specular reflectance is the Dirac-like response exclusively in the viewing direction that forms the same angle to the surface normal as the incident illumination. When viewed from the direction of specular reflectance, the surface normal is identical to the so-called *half angle* vector $h$ defined as the bisector of the angle formed by the rays of incident illumination and the viewing direction.

However, such mirror-like reflection is not realistic for most surfaces and the specular reflection is more often modelled as a lobe rather than as a Dirac pulse. The lobe is explained by the existence of micro-facets on any real surface that provide a range of orientations, a sort of standard deviation around the global surface point normal $n$. The range of introduced half-angles results in a lobe around the central axis of the strongest response arising from the global normal acting as the half-angle. There is a variety of models approximating the specular lobe e.g. the Phong [51] and Blinn-Phong [116] models compute a power of the dot product of respectively the specular reflectance vector+viewing direction and the normal+half angle vector. For synthetic data generation in the presented collocation error analysis, the Gaussian distribution, which the dot product formulations only approximate for larger dot product powers, is chosen because it is said to be a better model for the micro-facet-based specular reflection. The Gaussian distribution for specular reflectance modelling is formulated as:

$$f_{spec} = e^{-\left(\frac{\cos(h \cdot n)}{c_{rough}}\right)^2}$$

(4.25)

where $h$ is the half-angle direction as defined previously, $n$ is the surface normal and $c_{rough}$ is the roughness constant between 0 and 1 controlling the size of the specular lobe [117]. The constant models the degree of micro-facet presence i.e the relative surface smoothness and hence the normal fluctuation around the global surface normal $n$ arising from the set of micro-normal orientations.

The synthetic data generation procedure described is consistently used throughout Section 4.3.1 with any relevant parameter values specified along the way. Consistently with the real-life set-up in Figure 4.8a, the simulated light source location perturbation will be along the horizontal axis of the configuration in Figure 4.8b. Taking into account the set-up dimensions, the light source perturbation range of $[-700 mm, 700 mm]$ is considered meaningful, exhaustively covering both the typical miscollocations encountered and the extreme
Figure 4.10: Normal estimate error at point $x = [0, 0, 1200]$ with the perfectly perpendicular normal $n = [0, 0, -1]$ and Lambertian reflectance due to light source location perturbation (i.e. camera-light source miscollocation).

ones. Note that, also consistently with the real set-up, all three light sources are perturbed simultaneously.

1. **Collocation error analysis: normal estimate error.** In this part of the analysis, the depth of a surface point is assumed known, while the effect of the light source - camera miscollocation on the normal estimate is observed. The normals are computed by SVD decomposition of the matrix of three HS constraints. To introduce a light source location perturbation in the simulation, the intensities are generated according to Equation 4.23 with perturbed vectors $v_1$ and $v_2$ while the HS constraints are computed combining the expected sampling geometry with the corrupted intensities. The experiment is performed for a Lambertian surface point as well as one with the specular BRDF as discussed above.

**Lambertian surface point.** Consider a Lambertian point $x$ with the surface normal $n$, specifically: $x = [0, 0, 1200]$, $n = [0, 0, -1]$ and the Lambertian BRDF $f_r = f_{diff} = 1.0$.

For a point with a Lambertian reflectance model, perturbation only takes effect via the sampling geometry in Equation 4.23 as the BRDF $f_r$ is always constant and independent of the light source position.

The considered range of light source-camera miscollocations is swept in $10\ mm$ increments. Figure 4.10a shows the normal estimate error as a function of miscollocation in the range. For the most realistic miscollocation of $\pm 200\ mm$ the normal estimate error is about $16^\circ$.

The corresponding SVD residual, indicative of the quality of the normal estimate, is pre-
4.3. Evaluation

Presented in Figure 4.10b also as a function of miscollocation. As expected, the quality values in particular peak at 0 mm when the camera and the light source are in fact perfectly collocated. There is also no normal error in Figure 4.10a with the perfect collocation. Although the peak indicating estimate quality at 0 mm undoubtedly dominates in Figure 4.10b, there are also sporadic peaks of increased SVD residual with substantially perturbed light source locations. The observation shows that at this Lambertian point the SVD residual as a measure of estimate quality does not exactly match the trend of systematic increase shown by the normal error as a function of miscollocation.

Further the analysis is repeated for a plausible range of surface orientations, specifically by varying the orientation to be estimated within ±40° from the original perfectly perpendicular \( \mathbf{n} = [0, 0, -1] \). Note that, for simplicity, the surface normals, listed in the legend of Figure 4.11, are confined within the plane \( y = 0 \). Figure 4.11 reveals the dependence of the normal estimate error on the surface normal orientation. There are a few interesting observations to be made from the graph. Firstly, the error is lower when the normal is oriented in the direction opposite to the collocation perturbation. The error is only completely symmetrical with respect to the positive and negative perturbation for the original perfectly perpendicular surface orientation: \( \mathbf{n} = [0, 0, -1] \). Thirdly, even for a favourable surface orientation in the range considered, the normal estimate error never drops below...
Figure 4.12: Weighted specular lobe \((1 - \alpha)f_{\text{spec}}\). Parameters: roughness constant \(c_{\text{rough}} = 0.5\), weight \((1 - \alpha) = 0.4\).

\(10^\circ\) at the realistic perturbation of \(\pm 200\) mm.

**Specular surface point.** Now consider a specular point \(x\) with the surface normal \(n\), specifically: \(x = [0, 0, 1200]\), \(n = [0, 0, -1]\) and the specular BRDF \(f_r = \alpha f_{\text{diff}} + (1 - \alpha)f_{\text{spec}}\) where \(\alpha = 0.6\) and \(f_{\text{diff}} = 1.0\) as before. Unlike the Lambertian point, the measurement at a specular point \(x\) will be perturbed both via the sampling geometry and the BRDF, specifically its specular component \(f_{\text{spec}}\) as expressed in Equation 4.25 (recall that the half angle vector \(h\) is dependent on the knowledge of illumination direction as it by definition bisects the angle formed by the incident illumination ray and the viewing direction). The roughness parameter \(c_{\text{rough}}\) in Equation 4.25 is set to 0.5 to ensure a substantial specular component across the entire surface due to a wide specular lobe, as Figure 4.12 shows, mostly contained within \(-50^\circ \leq \angle(n \cdot h) \leq 50^\circ\).

The normal estimate error behaves very similarly as a function of miscollocation regardless of whether the sampled point is specular or Lambertian. In Figure 4.13a the two graphs showing the trend for the two respective models are indistinguishable except at the extremities of the light source perturbation range. Also, for a specular point, at the realistic miscollocation of \(\pm 200\) mm the normal error range as a function of surface normal variation at point \(x\) is still \(11^\circ - 17^\circ\), as for a Lambertian point in Figure 4.11.

The SVD residual behaves differently as the function of miscollocation for the two models. While the residual peaks unambiguously at 0 mm for both models with a value of the same order of magnitude \((10^{31})\), the magnitude of the SVD residual drops a lot more dramatically at non-zero miscollocations for the specular model \((10^3 \leq \text{SVD residual} \leq 10^4)\) than for the Lambertian model \((10^{16} \leq \text{SVD residual} \leq 10^{18})\). Also both the amplitude and fre-
4.3. Evaluation

Figure 4.13: Normal estimate error at point \( \mathbf{x} = [0, 0, 1200] \) with the perfectly perpendicular normal \( \mathbf{n} = [0, 0, -1] \) due to light source location perturbation (i.e. camera-light source miscollocation) for the Lambertian and specular reflectance models.

The frequency of the residual’s fluctuation as a function of miscollocation are substantially higher with the Lambertian model, which indicates penalty inconsistency. Any miscollocation is penalised more heavily and consistently with the specular model because the specular reflectance component \( f_{\text{spec}} \) is directly defined by the sampling geometry, which is unlike the constancy of the diffuse component \( f_{\text{diff}} \) in the Lambertian model. Note that, with the specular model, the light source location perturbation directly impacts the fundamental BRDF equality within the reciprocal pair through which the constraints are defined: the two geometry dependent BRDF samples involved in a reciprocal pair are independently corrupted by miscollocation of their respective light sources and are no longer equal. In contrast, with the Lambertian model, the fundamental equality is still valid given a miscollocation as the diffuse reflectance is a constant, independent of the sampling geometry, and the intensity measurements are corrupted via the foreshortening term in the image formation Equation 4.23 only.

2. Collocation error analysis: reconstruction from synthetic data. This second part of the collocation error analysis focuses on how the proposed reconstruction framework copes with the normal error due to miscollocation. To this end, reconstruction is performed on synthetic data with ground truth facilitating quantitative analysis of both depth and normal error. The study is performed on two synthetic objects, a pear (“Pear”) and a
sphere (“Sphere”), for a variety of reflectance models specifically one Lambertian and three specular each characterised by a different width of the Gaussian specular lobe $f_{\text{spec}}$. The reconstruction volume dimensions for Pear and Sphere are 95 mm $\times$ 107 mm $\times$ 100 mm and 160 mm $\times$ 180 mm $\times$ 100 mm respectively. Figure 4.14 shows examples of intensity images for each dataset whereas the defining parameters of the reflectance models used are presented in Table 4.1. The key difference between the three specular models is the size of the specular lobe defined via the roughness parameter $c_{\text{rough}}$.

Table 4.1: Reflectance models for synthetic data

<table>
<thead>
<tr>
<th>Reflectance Model</th>
<th>$f_{\text{diff}}$</th>
<th>$f_{\text{spec}}$ weight $(1 - \alpha)$</th>
<th>$c_{\text{rough}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lambertian</td>
<td>1</td>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>Specular 1</td>
<td>0.6</td>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>Specular 2</td>
<td>0.6</td>
<td>0.4</td>
<td>0.25</td>
</tr>
<tr>
<td>Specular 3</td>
<td>0.6</td>
<td>0.4</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Figure 4.14: Examples of input synthetic data for the collocation error analysis with varying reflectance models.

The synthetic set-up configuration generally follows the synthetic data generation procedure
Table 4.2: Additional parameters per dataset for the simulated CL HS set-up.

<table>
<thead>
<tr>
<th></th>
<th>Pear</th>
<th>Sphere</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{\text{cam}}$</td>
<td>500 mm</td>
<td>1000 mm</td>
</tr>
<tr>
<td>cameras directed at point</td>
<td>[10 mm, 70 mm, 1200 mm]</td>
<td>[0 mm, 0 mm, 1200 mm]</td>
</tr>
<tr>
<td>field of view</td>
<td>30°</td>
<td>30°</td>
</tr>
</tbody>
</table>

at the beginning of Section 4.3.1 in terms of the realistic dimensions ($d_{\text{scene}}$, $\angle R_1$, $\angle R_2$ etc.) assigned except for some minor alterations highlighted per dataset in Table 4.2 along with some additional configuration-related information. For example, for Pear the camera radius $r_{\text{cam}}$ needed to be reduced from 1000 mm to 500 mm to avoid excessive shadowing of image parts affecting reconstruction completeness, while the camera radius in Sphere reconstruction is closer to the dimensions of the real set-up in Figure 4.8a. In the present experiment for each dataset with each reflectance model considered (Table 4.1) five scenarios are tested:

1. ideal collocation of cameras and light sources
2. consistent\(^2\) miscollocation of 200 mm
   (the light source to the left of camera as in Figure 4.8a.)
3. consistent miscollocation of 500 mm
4. consistent miscollocation of 700 mm
5. consistent miscollocation of −200 mm (the light source to the right of camera)

The sets of graphs in Figure 4.15 show rms depth and normal error as a function of miscollocation for Pear and Sphere with the four different reflectance models. There are several conclusions that can be drawn from the quantitative results. Firstly, the proposed reconstruction framework (CtF MRF HS with dist.DNprior) is not particularly sensitive to the miscollocation introduced by the acquisition set-up design. At the miscollocation of 200 mm, most closely approximating the reality in Figure 4.8a, rms depth errors are between 1 mm and 1.5 mm for both datasets with all reflectance models. The corresponding

\(^2\)i.e. the same for all camera/light source pairs
Chapter 4. CL HS for dynamic scenes with uniform chromaticity

Figure 4.15: Quantitative comparison of rms depth and normal error as a function of miscollocation for Pear and Sphere with different reflectance models.
rms normal errors are between 5° and 7°, which is actually lower than the normal error range of 11° – 17° degrees estimated at surface point $x = [0, 0, 1200]$ in the first part of the collocation error analysis. These depth and normal error results validate the acquisition set-up as suitable for CL HS featuring CtF MRF HS with dist.DNprior as the reconstruction core.

As the specular lobe gets smaller becoming more insignificant (i.e. Specular 1 and 2), the specular model closely approaches, or even becomes indistinguishable from, the Lambertian model in the overall rms accuracy scores. In general, the results show that the performance as a function of miscollocation is more or less independent of the reflectance model with only small differences occurring at large miscollocations and not indicating any particular trend.

Figure 4.16 shows the effect of camera/light source miscollocation on the spatial distribution of depth and normal errors for two different objects with the Lambertian and Specular 3 reflectance models. The presence of miscollocation changes the spatial distribution pattern of both errors as well as their magnitude. The increase in depth and normal errors appears to concentrate in certain reconstruction regions: note how off-centre points are affected more than the ones more symmetrical with respect to the set-up. The result is consistent with the qualitative observation in 3D that through miscollocation the reconstructed mesh seems to rotate around the vertical axis. For realistic miscollocation such as 200 mm the rotation is however fairly small. Further, the depth and normal error distributions do not always follow the same pattern as, although a degree of depth-normal consistency is enforced via the dist.DNprior, there can still be coincidental similarity to the true normal at incorrect depths as well as the imperfections of the normal estimate at the correct depth.

For all datasets at the realistic miscollocation of 200 mm the errors lie under 4 mm depth-wise and under 10° in the normals. On visual inspection of the spatial error distributions, the specular model often seems to have lower errors for off-centre surface points, affecting smaller spatial regions with a given intensity, than the Lambertian model. However, especially for the realistic miscollocation of 200 mm, the advantage is neither substantial, in terms of magnitude or spatial extent of improvement, nor universally consistent for all metrics and datasets (see the depth error distributions of Sphere). The observation supports the conclusion stated earlier based on the rms trends that the performance of CtF MRF HS
Figure 4.16: Spatial distribution of depth and normal errors for Pear and Sphere with Lambertian and Specular 3 reflectance models for different miscollocation scenarios.
with dist.DNprior as a function of miscollocation is overall independent of the reflectance model.

From the collocation error analysis we conclude that the proposed algorithm is well equipped for dealing with the error introduced through miscollocation. At the realistic light source location perturbation of 200 mm, the perceived 3D mesh rotation, the local as well as the rms depth/normal errors are all at acceptable levels.

### 4.3.2 Photometric calibration results

In this part of evaluation the estimated spatial distributions of the photometric parameter $\mu$ for the multi-spectral Helmholtz cameras of the real set-up in Figure 4.8a are presented and discussed. The spatial distributions are presented as heat maps (photometric maps) in Figure 4.17 with the left-hand and right-hand columns respectively corresponding to the set-up configured for the specific static (Section 4.3.3) and dynamic (Section 4.3.4) scenes to be reconstructed.

The photometric maps are shown for each pair of reciprocal multi-spectral Helmholtz cameras in the set-up. The distributions within each such pair are mutually consistent and are therefore presented on the same colour scale as shown. There can be a scaling factor inconsistency of coefficients between the pairs, which however does not affect the mutual consistency of the intensity constraints computed from the pairs, dependent only on the constraint vector orientation not its magnitude. There is a visible spatial variation of $\mu$ in all photometric maps but the values seem to be confined in the range between 0.6 and 1.6.

In each pair one camera is calibrated directly and the parameters are transferred onto its reciprocal partner as described in Section 4.2.3.

Some cameras, such as $R_2$ in the dynamic scene calibration set, act as the directly calibrated camera in multiple pairs because transferring from (rather than to) them gives the greatest combined spatial coverage for the pair. Given the framing of the static and dynamic scenes respectively, the maps presented in Figure 4.17 provide enough spatial coverage to reconstruct the foreground scene within the visual hull formed by the three cameras in the configuration. Depending on the depth hypothesis scope considered, some outlier background voxels in the reconstruction volume may be sampling the uncalibrated spatial extremities of the photometric maps. Since it only affects the background voxels, any sim-
ple spatial extrapolation method, or even substitution of a constant, where \( \mu \) is unavailable would be a sufficient approach. The provision promotes completeness of the reconstruction as more surface points on the outskirts of the 2.5D scope can be kept in the reconstruction scope instead of being eliminated due their incompletely defined depth hypothesis scope (i.e. background voxels without photometric calibration).

\[
\begin{array}{c|c|c|c|c}
R_1 & R_2 & R_2 & R_3 & R_2 \\
R_1 & R_2 & R_2 & R_3 & R_2 \\
R_2 & R_3 & R_1 & R_3 & R_1 \\
\end{array}
\]

Figure 4.17: Heat map representation of the spatial photometric parameter \( \mu \) distributions for each pair of reciprocal multi-spectral Helmholtz cameras in the set-ups of the static (left-hand side) and dynamic (right-hand side) scene capture.

### 4.3.3 Static scenes

Evaluation with static scenes provides an opportunity to gauge the accuracy of the proposed CL HS quantitatively and qualitatively under more controlled conditions than in dynamic scene reconstruction. In this chapter, only objects with uniform chromaticity are considered for reconstruction. Specifically for this evaluation, the greyscale chromaticity (and coincidently white colour) is chosen as it matches the chromaticity of the calibration plane used during photometric reconstruction.
The evaluation on static scenes consists of two parts. Firstly, for quantitative evaluation a portion of the white matte photometric calibration plane (“Plane”) is reconstructed. Secondly, the qualitative evaluation of accuracy is performed on a white specular mug (“White Mug”). In this section, as in the rest of evaluation for this chapter, Coarse-to-fine Bayesian Helmholtz Stereopsis solved by MRF optimisation (CtF MRF HS) implies the use of the distance-based depth-normal consistency prior (dist.DNprior) enforcing integrability. Standard ML HS embedded into the coarse-to-fine framework is referred to as Coarse-to-fine Maximum Likelihood Helmholtz Stereopsis (CtF ML HS).

As in Chapter 3, the results are rendered with both flat and smooth shading as appropriate to showcase either the pure geometry or the overall quality with both depth and normal estimates considered. Also as before a default fixed reflectance model is used for result rendering as the true surface BRDF estimation is beyond the scope of this work targeting geometry reconstruction regardless of its reflectance model.

**Input intensity images**

Figure 4.18: Input intensity images for Plane reconstruction.

**Quantitative evaluation: Plane.** The plane used in this experiment is the calibration board in one of its positions captured for photometric calibration. The test plane thus chosen from the photometric calibration data pool did not belong to the set actually used to obtain the photometric calibration maps. The markers present on the calibration board allow to compute the ground truth plane orientation through triangulation. Having the entire plane including the black markers in the reconstruction is undesirable because, while black also belongs to the greyscale chromaticity family, the brightness of the colour approaches zero resulting in little or no signal on any of the channels and a low signal-to-noise ratio. The low intensity in the marker region would taint the quantitative evaluation, potentially affecting the reconstruction’s global as well as local shape. Equally, the global shape could be distorted if the reconstruction were attempted on the entire plane with holes for the
markers. So instead reconstruction is performed on one uniformly white central patch of the calibration plane within the area delimited by the markers.

The test plane in the three camera views is shown in Figure 4.18. The reflectance model of the object is approximately Lambertian. The complexity of Plane is in the geometric location ambiguity as, due to flatness, sets of depths with similar plausibility scores and normals are generated. Equally any non-flatness of the calibration surface (e.g. due to the paper rising up from the board) would also introduce deviation from the triangulated ground truth (this source of error was also indicated by Jankó et al. in their evaluation of photometric calibration for WL HS in [106]). In selecting the test segment of the plane attempts were made to avoid even minimal non-flatness due to air bubbles under the paper surface as the proposed coarse-to-fine reconstruction framework produces highly precise reconstructions.

Plane was reconstructed in a coarse-to-fine framework consisting of 5 iterations from the initial resolution of 1 mm/0.5 mm (spatially/depthwise) and the spatial subdivision termination threshold at 0.25 mm. The heat maps illustrating the spatial distribution of depth and normal error (Figure 4.19) are instructive comparing the quality of the patch reconstruction using three methods: 1. CtF ML HS with photometric calibration (of Helmholtz cameras) (PhotoCalib); 2. CtF MRF HS without PhotoCalib and 3. CtF MRF HS with PhotoCalib. The overall quantitative metrics comparing reconstructions to the ground truth are presented in Table 4.3. The rms depth and normal errors are computed at the 100% Middlebury accuracy threshold (i.e. no outlier elimination). Calibrated CtF MRF HS has
proven itself capable of sub-millimeter depth reconstruction accuracy. The accuracy drops by half a millimetre if photometric calibration is omitted. Remarkable is the deterioration in the quality of the normals without photometric calibration. The rms normal error of CtF MRF HS without PhotoCalib is even significantly worse than that of calibrated CtF ML HS. In terms of depth the quality of CtF ML HS is unacceptable. The conclusion that can be drawn from the observation is that the choice of the reconstruction principles (ML versus Bayesian) directly affects depth estimates whereas the presence or absence of photometric calibration is paramount for the accuracy of normal estimation.

Table 4.3: Plane reconstruction: a quantitative evaluation.

<table>
<thead>
<tr>
<th></th>
<th>rms depth error [mm]</th>
<th>rms normal error [°]</th>
</tr>
</thead>
<tbody>
<tr>
<td>CtF ML HS with PhotoCalib</td>
<td>32.922</td>
<td></td>
</tr>
<tr>
<td>CtF MRF HS without PhotoCalib</td>
<td>1.5701</td>
<td></td>
</tr>
<tr>
<td>CtF MRF HS with PhotoCalib</td>
<td>0.99317</td>
<td></td>
</tr>
<tr>
<td>CtF ML HS with PhotoCalib</td>
<td>3.0812</td>
<td></td>
</tr>
<tr>
<td>CtF MRF HS w/o PhotoCalib</td>
<td>15.259</td>
<td></td>
</tr>
<tr>
<td>CtF MRF HS with PhotoCalib</td>
<td>2.5843</td>
<td></td>
</tr>
</tbody>
</table>

Qualitative evaluation: White Mug. The object is characterised by a more complex, specular model with pronounced multi-chromatic specularities. The reconstructions presented for qualitative evaluation are obtained with CtF MRF HS consisting of 5 iterations with a fine initial resolution of 1 mm/0.25 mm (spatially/depthwise). The automatically detected spatial subdivision termination threshold is 0.5 mm. The point cloud is meshed without explicit integration using the method based on known geometric ordering of vertices without explicit integration (NoInt) from Chapter 3. Several important qualitative observations can be made from Figure 4.20. Firstly, the global geometric accuracy achieved by CL HS with just three reciprocal pairs using the tailored reconstruction CtF MRF HS framework is comparable to the results obtained in Chapter 3 by WL HS with eight reciprocal pairs and the same reconstruction method. This observation goes to highlight the impressive accuracy of CL HS with CtF MRF HS with the input information drastically reduced to the minimum of three reciprocal pairs. Secondly, the White Mug reconstruction clearly indicates the importance of photometric calibration: both views shown reveal a global curvature deformation of the mug in the absence of photometric calibration. The results presented in smooth shading indicate that with photometric calibration the algorithm recovers better normals in the region of specular
Chapter 4. CL HS for dynamic scenes with uniform chromaticity

Figure 4.20: Qualitative evaluation of CL HS on the uniformly greyscale static scene White Mug: a comparison of calibrated CtF ML HS against CtF MRF HS with and without photometric calibration.

Table 4.1: Input intensity images and reconstruction results for the uniformly greyscale static scene White Mug.

<table>
<thead>
<tr>
<th>CtF ML HS FullCalib</th>
<th>CtF MRF HS no PhotoCalib</th>
<th>CtF MRF HS FullCalib</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Input intensity images" /></td>
<td><img src="image2" alt="Reconstruction results, flat shading" /></td>
<td><img src="image3" alt="Reconstruction results, smooth shading" /></td>
</tr>
</tbody>
</table>

Highlights making them much less noticeable in the mesh. Thirdly, confirming the findings in Chapter 3, standard ML HS, even with the coarse-to-fine extension, is shown to be completely inadequate as a reconstruction algorithm for CL HS.

The quantitative and qualitative evaluation of the proposed algorithm on static scenes with greyscale chromaticity has demonstrated accuracy in the face of both the geometric ambiguity of Plane and the photometric complexity of White Mug. The necessity of photometric ...
calibration has also been validated on both examples. In the following section, evaluation results for reconstruction of non-rigid dynamic scenes, also with greyscale chromaticity, will be presented to validate CL HS for the targeted application.

4.3.4 Dynamic scenes

Having verified the accuracy of the proposed CL HS on static scenes, let us present the results for two dynamic datasets featuring non-rigid deformation of uniformly white materials, specifically a white laminated sheet (“WLS”) and the cloth of a blouse (“Blouse”). Despite their uniformly greyscale chromaticity (specifically approximately white colour), the reflectance models of both materials, particular the laminated paper, are complex. WLS is characterised by extreme specular highlights clearly prohibitive of reliable disparity estimation for reconstruction with Conventional Stereo. The reflectance model is too complex to estimate reliably for reconstruction with Photometric Stereo. Blouse, although somewhat glossy in appearance, has a simpler reflectance model but a much more complex deformation geometry with its numerous material folds and creases as well as the greater freedom of deformation due to the complete non-rigidity of cloth. In addition, the material of Blouse is not fully opaque which will corrupt the generated Helmholtz constraints as Helmholtz reciprocity is violated through the partial illumination loss due to transparency (for further discussion see Chapter 5 Section 5.3). Furthermore, both WLS and Blouse are virtually textureless which would have posed difficulties for Conventional Stereo in establishing intensity matches even if the reflectance model had been Lambertian.

In this section performance of the proposed CL HS with CtF MRF HS as the reconstruction core, as before featuring the tailored dist.DNprior, is validated on the challenging WLS and Blouse datasets. The input video with the material deformation consists of 201 frames at 25 frames per second. The dynamic sequences are reconstructed frame-by-frame with each frame processed independently. At the high resolutions facilitated by the coarse-to-fine framework the surface can be obtained from the point cloud without explicit integration as in Section 4.3.3. Other reconstruction parameters are specified in Table 4.4.

Figures 4.21 and 4.22 for WLS and Blouse respectively each show a sample set of five frame reconstructions, 25 frames apart. Consistently with the static scene results, the reconstruction cores of CL HS compared are: 1. CtF ML HS with PhotoCalib; 2. CtF MRF HS
Table 4.4: CL HS evaluation: reconstruction parameters for WLS and Blouse.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial resolution spatial:</td>
<td>3 mm</td>
</tr>
<tr>
<td>Initial resolution depthwise:</td>
<td>0.5 mm</td>
</tr>
<tr>
<td>Reconstruction volume Blouse and WLS:</td>
<td>$120 \text{ mm} \times 10^7 \text{ mm} \times 30^7 \text{ mm}$</td>
</tr>
<tr>
<td>Spatial sub-division ceases at:</td>
<td>0.375 mm</td>
</tr>
<tr>
<td>Number of coarse-to-fine iterations:</td>
<td>5</td>
</tr>
<tr>
<td>Number of MRF optimisation cycles per iteration:</td>
<td>6</td>
</tr>
<tr>
<td>Prior used for CtF MRF HS:</td>
<td>dist.DNprior</td>
</tr>
<tr>
<td>Surface integration method:</td>
<td>NoInt</td>
</tr>
</tbody>
</table>

without PhotoCalib and 3. CtF MRF HS with PhotoCalib. The purpose of the experiment is to validate the choice of CtF MRF HS over CtF ML HS as well as the necessity of PhotoCalib. The results are presented with flat shading to showcase the accuracy of raw geometry obtained, unmasked by shading using photometric normals.

CtF MRF HS produces results with a remarkable level of detail resolution (see Blouse) and global accuracy. Due to the method’s reflectance model independence, even the most drastic non-Lambertian behaviour (e.g. the specularites of WLS) and unconstrained geometry are successfully coped with reproducing the folds, creases and domes of the deforming objects in the reconstructed meshes. CtF MRF HS, both with and without PhotoCalib, as before clearly outperforms the fully calibrated CtF ML HS. Photometric calibration is also clearly essential for global accuracy of the reconstructed material patch as can be observed comparing the performance of CtF MRF HS with and without photometric calibration on both Blouse and WLS. The characteristic global shape errors without photometric calibration include the somewhat retracted material patch orientation and the oddly enlarged and/or stretched prominent features. The uncalibrated method also seems to resolve the creases of Blouse less well. In frames 76 and 101 of the WLS sequence saturated regions result in reconstruction of pseudo-geometries, flat crease-like folds, on the surface. The artefact is present to some extent in both calibrated and uncalibrated reconstruction (the photometric calibration does not account for sensor saturations), yet is significantly more apparent in the uncalibrated result due to its erroneous overall patch orientation. Upon close examination of the videos with full dynamic sequence reconstructions, CtF MRF HS shows similar temporal stability with and without calibration. A frame-by-frame comparison however reveals a consistent blurring of fine structural detail when the method is uncalibrated.
4.3 Evaluation

Figure 4.21: Reconstruction of complex dynamic scenes with uniform chromaticity using CL HS: WLS. Input: intensity images of $C_1$, $C_2$, and $C_3$. Reconstruction cores compared: calibrated CtF ML HS, uncalibrated CtF MRF HS and calibrated CtF MRF HS (proposed); Initial sampling resolution: $3\, mm/0.5\, mm$ (spatially/depthwise); Integration: NoInt for all. Note that only a portion of the scene was reconstructed due to framing.
Chapter 4. CL HS for dynamic scenes with uniform chromaticity

Figure 4.22: Reconstruction of complex dynamic scenes with uniform chromaticity using CL HS. Blouse. Input: intensity images of $C_1$, $C_2$, and $C_3$. Reconstruction cores compared: calibrated CtF ML HS, uncalibrated CtF MRF HS and calibrated CtF MRF HS (proposed). Initial sampling resolution: $3\text{ mm}/0.5\text{ mm}$ (spatially/depthwise); Integration: NoInt for all. Note that only a portion of the scene was reconstructed due to framing.

<table>
<thead>
<tr>
<th></th>
<th>CtF MRF HS with PhotoCalib</th>
<th>CtF MRF HS</th>
<th>CtF ML HS with PhotoCalib</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame 1</td>
<td><img src="image1" alt="CtF MRF HS with PhotoCalib" /></td>
<td><img src="image2" alt="CtF MRF HS" /></td>
<td><img src="image3" alt="CtF ML HS with PhotoCalib" /></td>
</tr>
<tr>
<td>Frame 26</td>
<td><img src="image1" alt="CtF MRF HS with PhotoCalib" /></td>
<td><img src="image2" alt="CtF MRF HS" /></td>
<td><img src="image3" alt="CtF ML HS with PhotoCalib" /></td>
</tr>
<tr>
<td>Frame 51</td>
<td><img src="image1" alt="CtF MRF HS with PhotoCalib" /></td>
<td><img src="image2" alt="CtF MRF HS" /></td>
<td><img src="image3" alt="CtF ML HS with PhotoCalib" /></td>
</tr>
<tr>
<td>Frame 76</td>
<td><img src="image1" alt="CtF MRF HS with PhotoCalib" /></td>
<td><img src="image2" alt="CtF MRF HS" /></td>
<td><img src="image3" alt="CtF ML HS with PhotoCalib" /></td>
</tr>
<tr>
<td>Frame 101</td>
<td><img src="image1" alt="CtF MRF HS with PhotoCalib" /></td>
<td><img src="image2" alt="CtF MRF HS" /></td>
<td><img src="image3" alt="CtF ML HS with PhotoCalib" /></td>
</tr>
</tbody>
</table>
Accurate 3D dynamic scene reconstruction given unconstrained reflectance behaviour is an all-around challenging task never previously tackled in its full complexity. In this chapter, a solution has been proposed based on the Bayesian reformulation of reflectance-model-independent Helmholtz Stereopsis (HS) from Chapter 3. The proposed solution, the multi-spectral Colour Helmholtz Stereopsis (CL HS), fundamentally alters the pipeline of HS from acquisition to integration for the first time generalising HS to dynamic scenes. The new acquisition set-up based on wavelength-multiplexing has been shown to introduce additional challenges related to ensuring inter-channel signal consistency. In response to the new challenges a practical photometric calibration procedure has been incorporated into the framework. The procedure has been validated as essential in the reconstruction pipeline by comparison to the results of the uncalibrated pipeline for a variety of real complex objects. The proposed calibration procedure generalises the method in [106] for WL HS to the multi-spectral CL HS addressing the issues of calibration in a multi-Helmholtz-camera-pair configuration with wavelength multiplexing, practicality of acquisition and scalability.

A thorough evaluation of the proposed novel acquisition set-up has been carried out. Specifically, the error due to the approximate camera/light source collocation has been studied using synthetic data. At realistic miscollocations the observed rms depth error of $1 - 1.5 \, mm$ and the rms normal error of $5^\circ - 7^\circ$, as well as the visual appearance of the realistically perturbed results relative to the unperturbed ones, have been found acceptable validating the proposed set-up.

The initial version of CL HS proposed in this chapter targets scenes with uniform chromaticity but otherwise no further constraints are placed on the reflectance behaviour. The chapter evaluates the accuracy quantitatively on a real planar surface obtaining sub-millimetre rms depth accuracies and the normal rms accuracy of roughly $2 - 3^\circ$. A highly specular static mug reconstruction showcases the method’s performance on curvature revealing a highly plausible global shape of the object. Finally, reconstruction of the two dynamic datasets has covered validation of accuracy in the face of both geometric intricacy and photometric complexity while capturing a temporally consistent sequence. Robustness to extreme photometric complexity has been put to the test by the white laminated sheet throwing sporadic unpredictable specular patterns. The ability to resolve intricate geometries has
been evaluated by the blouse cloth deformation dataset with its utter non-rigidity and the numerous creases and folds. The proposed pipeline for CL HS has been validated with regards to the claims on robustness to geometric and photometric complexity as well as temporal consistency. The occasional local pseudo-geometries in the regions of (near-) sensor saturation (where intensity information is missing or inaccurate) are fairly muted and do not corrupt the global shape.

The presented high-resolution reconstructions of material deformation with uniform chromaticity place CL HS on the map as a competitive dynamic reconstruction approach for scenes of high photometric complexity. There are however limitations to the approach proposed in this chapter. Firstly, the uniform chromaticity of the scene must match the reference chromaticity defined by the chromatic properties of the plane used in the photometric calibration procedure. Unless the exact same material is used for both calibration and reconstruction, there is no guarantee, without actually measuring it, that the test object has exactly the same chromatic properties as the reference. Reference chromaticity approximation in choosing appropriate test objects by visual inspection will have introduced errors, although in practice it seems they are not prohibitive. The second limitation is the inability of the proposed pipeline to handle spatially varying chromaticity as it must uniformly match the single reference. Both these limitations will be addressed in the next chapter removing the remaining chromaticity-related reflectance constraints. The extension to scenes with arbitrary spatially-varying chromaticity will allow CL HS to reach its full potential in accurate dynamic scene reconstruction totally independent of the reflectance properties.
Chapter 5

Colour Helmholtz Stereopsis for Dynamic Scenes with Arbitrary Spatially-varying Chromaticity
In Chapter 4 a novel approach of Colour Helmholtz Stereopsis (CL HS) was introduced that for the first time permitted reconstruction of dynamic scenes with arbitrary \textit{a priori} unknown reflectance properties. The method showed high accuracy of reconstruction on both static and dynamic scenes with an input set of just three intensity images acquired instantaneously. The only reflectance limitation of the method proposed in Chapter 4 was the requirement of uniform chromaticity matching (or at least approximating) the reference chromaticity.

In this chapter the requirement of uniform reference chromaticity is removed as a generalised pipeline for CL HS is proposed. The extended pipeline is able to reconstruct dynamic scenes with arbitrary spatially-varying chromaticity making the reflectance model of the target scene truly unconstrained. The extended functionality of generalised CL HS is facilitated by a novel chromaticity calibration procedure essentially reconciling the reference chromaticity, relative to which equipment is calibrated, to the actual chromaticity of the reconstructed surface. Only restricting the relationship between the reference and the scene chromaticities to equality, as in Chapter 4, allows to by-pass such a procedure. Hence, the proposed chromaticity calibration greatly expands the applicability scope of CL HS.

The chapter begins in Section 5.1.1 with the methodology for the generalised CL HS where chromaticity is incorporated into the core constraint equations from Chapter 4 alongside photometric calibration. The notion of chromaticity, both in physical terms and in the pipeline of CL HS, is subsequently discussed in Section 5.1.2. The proposed spatio-temporal chromaticity calibration procedure introduced in Section 5.2 consists of 1. static estimation of the spatial chromaticity distribution and 2. propagation of the estimate throughout the sequence. Section 5.3 presents a discussion of complex reflectance behaviour, such as sub-surface scattering/translucency and transparency, as a source of error for both reconstruction and chromaticity calibration procedures. The generalised pipeline of CL HS is summarised in Section 5.4 together with the more practical considerations regarding operation of the algorithm. In particular, the influence of inter-channel cross-talk in a CL HS reconstruction configuration of Helmholtz cameras is addressed.

The evaluation in Section 5.5 of the chapter starts off by looking at the estimated scene chromaticity distributions ranging from uniform to spatially-varying. The reconstruction algorithm itself is evaluated on both static and dynamic scenes. The evaluation aims to
demonstrate the importance of full calibration for scenes with unconstrained chromaticity. The dataset pool contains complex scenes with chromaticity deviating radically from the reference in a spatially non-uniform manner but also some examples resembling the ones presented in Chapter 4 with the chromaticity approximating the reference. The former datasets challenge the pipeline to the fullest testing the efficiency of the essential in this case chromaticity estimation. The latter (near-reference chromaticity) examples are presented in order to establish whether explicitly calibrating for chromaticity results in better reconstructions for these datasets. The evaluation illustrates that depending on the complexity of chromaticity distribution with respect to the reference the need for calibration may vary from the exact per-pixel to a spatial-average-based or even to none at all. Due to the Bayesian formulation of HS from Chapter 3, the results show a degree of robustness to signal corruption by complex sub-surface reflectance phenomena, which are strictly speaking not in the scope of HS. In order to illustrate the effective limitations of the proposed generalised CL HS, in Section 5.5.4 particularly challenging datasets are discussed together with the underlying reasons for their complexity which in practice render them out of the technique’s scope. The scope of evaluation allows one to draw conclusions on the role and performance of the proposed chromaticity calibration as well as the prowess of the generalised CL HS as a whole as a reconstruction algorithm for scenes with arbitrary spatially-varying reflectance properties.

5.1 Colour Helmholtz Stereopsis with arbitrary chromaticity

Colour Helmholtz Stereopsis (CL HS) introduced in Chapter 4 uses multi-spectral illumination to obtain intensity samples needed to compute HS constraints. The reflectance and absorption properties of a surface point, encapsulated in its chromaticity, play a role in determining the observed intensities together with the photometric characteristics of the acquisition equipment and the sampling geometry.
5.1.1 Formalisation

Let us for convenience recapitulate the image formation equation for CL HS from Chapter 4:

\[ i_{1,2} = \rho_{2}(v_{2})\sigma_{1}(v_{1})p_{1,2}f_{d}(v_{2}, v_{1}) \frac{v_{2} \cdot n}{\|c_{2} - x\|^2} \]  \hspace{1cm} (5.1)

where surface point \( x \) with normal \( n \) is sampled. The point has the chromaticity coefficient \( p_{1,2} \) when illuminated by light source \( S_{2} \) with the spatial radiance distribution \( \rho_{2} \) and viewed by camera \( C_{1} \) with the spatial sensor sensitivity \( \sigma_{1} \). Further the point is characterised by the directional BRDF component \( f_{d}(v_{2}, v_{1}) \) independent of the sampling wavelength. In conjunction with the analogous \( i_{2,1} \), Equation 5.1 gives rise to the HS constraint:

\[ \left( \frac{\mu_{1}(v_{1})i_{1,2}}{p_{1,2}\|c_{1} - x\|^2} v_{1} - \frac{\mu_{2}(v_{2})i_{2,1}}{p_{2,1}\|c_{2} - x\|^2} v_{2} \right) \cdot n = 0 \]  \hspace{1cm} (5.2)

The previous chapter targets scenes with uniform reference chromaticity for which the set-up is \textit{a priori} calibrated by choosing a calibration target with certain chromatic characteristics: \( p_{1,2} = p_{1,2}^{ref} \) and \( p_{2,1} = p_{2,1}^{ref} \). The corresponding normal constraint for that formulation is:

\[ \left( \frac{\mu_{1}(v_{1})i_{1,2}}{\|c_{1} - x\|^2} v_{1} - \frac{\mu_{2}(v_{2})i_{2,1}}{\|c_{2} - x\|^2} v_{2} \right) \cdot n = 0 \]  \hspace{1cm} (5.3)

where \( \mu_{1}' = \frac{\mu_{1}}{p_{1,2}} \) and \( \mu_{2}' = \frac{\mu_{2}}{p_{2,1}} \) are the relative photometric parameter distributions with respect to some chosen reference expressed by the chromaticity coefficients \( p_{1,2}^{ref} \) and \( p_{2,1}^{ref} \). The absolute photometric distributions \( \mu_{1} \) and \( \mu_{2} \) are not accessible as the reference parameters \( p_{1,2}^{ref} \) and \( p_{2,1}^{ref} \) are not known. Let us therefore instead express \( \mu_{1} \) and \( \mu_{2} \) in Equation 5.2 as functions of the available \( \mu_{1}' \) and \( \mu_{2}' \) respectively:

\[ \left( \frac{p_{1,2}^{ref}\mu_{1}'(v_{1})i_{1,2}}{p_{1,2}\|c_{1} - x\|^2} v_{1} - \frac{p_{2,1}^{ref}\mu_{2}'(v_{2})i_{2,1}}{p_{2,1}\|c_{2} - x\|^2} v_{2} \right) \cdot n = 0 \]  \hspace{1cm} (5.4)

since \( \mu_{1} = p_{1,2}^{ref}\mu_{1}' \) and \( \mu_{2} = p_{2,1}^{ref}\mu_{2}' \). If the chromaticity of the reconstructed surface matches the reference as in the previous chapter i.e. \( p_{1,2} = p_{1,2}^{ref} \) and \( p_{2,1} = p_{2,1}^{ref} \), Equation 5.4 reduces to the special case constraint in Equation 5.3. If the chromaticity is not the reference, the chromatic discrepancy relative to the reference of photometric calibration is expressed via
the relative chromaticity coefficients \( p'_{2,1} = \frac{p_{2,1}}{p'_{2,1}} \) and \( p'_{1,2} = \frac{p_{1,2}}{p'_{1,2}} \). The normal constraint can thus be re-written as:

\[
\left( \frac{\mu'_1(v_1)\delta_{1,2}}{p'_{1,2} ||c_1 - x||^2} v_1 - \frac{\mu'_2(v_2)\delta_{2,1}}{p'_{2,1} ||c_2 - x||^2} v_2 \right) \cdot n = 0. \tag{5.5}
\]

In the previous chapter the scope is limited to surfaces of uniform chromaticity that matches the reference, which results in the relative chromaticity of unity. In that case, the photometric calibration of the equipment already encapsulates the chromaticity properties of the reconstructed surface as the latter are identical to those of the reference chromaticity. In this chapter, the chromaticity of the reconstructed surface is allowed to be arbitrary and spatially-varying necessitating its estimation. The procedure derived in Section 5.2.1 estimates relative per-pixel chromaticity as defined above using the same reference chromaticity as in the procedure for photometric equipment calibration discussed at length in Chapter 4. Note that this reference correspondence results in the system’s overall independence of any reference: the constraint in Equation 5.5 reduces to the original constraint in Equation 5.2 formulated in terms of the absolute values as the reference dependence in the photometric distribution of \( \mu' \) cancels against the same dependence in the relative chromaticity estimate \( p' \).

Let us now briefly delve into the physical meaning of surface chromaticity for better comprehension of the role and behaviour of the parameter of relative chromaticity \( p' \) in the proposed methodology of generalised CL HS.

### 5.1.2 The notion of chromaticity

Chromaticity can be defined in relation to the tristimulus values \( R, G \) and \( B \) which are respectively the total red, green and blue signal obtained by integrating the corresponding continuous red, green and blue responses as a function of wavelength curves (i.e. \( r(\lambda), g(\lambda) \))
and \( b(\lambda) \) respectively) weighted by the spectral power distribution function \( P(\lambda) \) [118]:

\[
\begin{align*}
R &= \int_0^\infty P(\lambda)r(\lambda)\delta\lambda \\
G &= \int_0^\infty P(\lambda)g(\lambda)\delta\lambda \\
B &= \int_0^\infty P(\lambda)b(\lambda)\delta\lambda
\end{align*}
\] (5.6)

The chromaticity triplet \([p_r, p_g, p_b]\) is then defined as the tristimulus triplet normalised by the sum of the three values:

\[
\begin{align*}
p_r &= \frac{R}{R + G + B} \\
p_g &= \frac{G}{R + G + B} \\
p_b &= \frac{B}{R + G + B}
\end{align*}
\] (5.7)

From Equation 5.7 it is clear that chromaticity describes only the relative inter-channel relationship, not the absolute intensity values. In other words, relating chromaticity to the Hue Saturation Brightness (HSB) colour model, one can say that chromaticity unambiguously describes the hue and the saturation of the colour but not its brightness. In the HSB model, hue is the dominant identity of the colour within the colour space. In the continuous response as a function of wavelength curve, hue can be identified as the wavelength within the white spectrum with the strongest response. Saturation is a measure of colour purity which can be observed by the width of the measured spectrum around the hue peak. The wider the spectrum, the more wavelengths will have a substantial response essentially “washing out” the purity of the dominant wavelength of the hue. Finally, brightness relates to the absolute strength of the integrated tristimulus components from Equation 5.6. Clearly, due to normalisation by the sum of the tristimulus components in Equation 5.7, chromaticity is independent of brightness.

The direct implication of chromaticity’s independence of brightness is the many-to-one nature of the function mapping colours onto chromaticity values. All colours with the same inter-channel relationship will map onto the same chromaticity value. The concept is best illustrated by the white-greyscale equivalence: all greyscale values from black ([0, 0, 0]) to white ([255, 255, 255]) are characterised by the relationship of equal response on all
three channels and therefore will map onto the same chromaticity triplet, for instance to \(\left[\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right]\) in case of normalisation of the chromaticity triplet to unity. More generally, any set of colours characterised by the same inter-channel relationship cannot be distinguished based on the chromaticity information. For the CL HS normal constraint in Equation 5.2 only the inter-channel ratio matters because any scaling factor of the chromaticity triplet is irrelevant in the homogeneous expression.

In Equation 5.2 a chromaticity coefficient is denoted by \(p_{c,l}\) where \(c\) and \(l\) refer respectively to the camera and the light source that generate intensity sample \(i_{c,l}\). The three sources of the set-up, i.e. \(l = [R, G, B]\), have red \(r(\lambda)\), green \(g(\lambda)\) and blue \(b(\lambda)\) spectra and hence the estimated chromaticity triplet is in direct correspondence to the theoretical definition of chromaticity in Equation 5.7. It must be said however that the red, green and blue spectra of the equipment relative to which the chromaticity is estimated only approximate the absolute tristimulus spectra. The approximation does not introduce any reconstruction errors as the same spectra are used for sampling in the reconstruction procedure. Similarly, in Section 5.2.1 it will be discussed how instead of the absolute chromaticity \([p_R, p_G, p_B]\) derived here, an estimate relative to the reference chromaticity \([p'_R, p'_G, p'_B]\) = \([\frac{p_R}{p_{ref_R}}, \frac{p_G}{p_{ref_G}}, \frac{p_B}{p_{ref_B}}]\) is recovered. This relativism is intentional tailoring the chromaticity estimate to the rest of the CL HS system and making the whole pipeline independent of the reference chromaticity.

The absolute chromaticity triplet is an inherent property of the surface point independent of the acquisition equipment. It has been described how relative chromaticity must be estimated for CL HS to provide compatibility between different parts of the system. In this work, for convenience of relating to intensity samples, the spatial distribution of the relative chromaticity triplet (i.e. the chromaticity map) is acquired per image pixel for each camera \(c = \{1, 2, 3\}\) independently. Not every surface is characterised by the same level of complexity in the chromaticity distribution. It is more straightforward to deal with surfaces of uniform chromaticity, regardless of its relation to the reference, because in this case an average chromaticity triplet valid for the entire surface would suffice as chromaticity calibration. The most difficult case is that of a spatially-varying chromaticity distribution where the chromaticity triplet must be obtained per-pixel. Per-pixel mode of chromaticity calibration is challenging particularly in dynamic scene reconstruction where chromaticity maps evolve in correspondence to the changing intensity images. Section 5.2
proposes a spatio-temporal surface chromaticity calibration procedure consisting of two alternative modes of chromaticity estimation (average and per-pixel) as well as an additional chromaticity propagation stage to facilitate reconstruction of dynamic scenes. Further it should be remarked that given a single reference chromaticity and the same spectral camera sensor characteristics relative to the colour filter spectra, the estimated chromaticity should be consistent between camera views. However, the theory of chromaticity calibration, as well as of reconstruction by HS, only holds with the assumption of the BRDF reflectance model (i.e. direct surface reflectance without sub-surface transport, see Figure 5.5). Section 5.3 will discuss how chromaticity estimation is affected by producing camera-to-camera inconsistencies when the surface deviates from the assumption.

5.1.3 Overview of the pipeline

A schematic representation of the key stages of CL HS for surfaces with arbitrary spatially-varying chromaticity is presented in Figure 5.1 for a global overview. Compared to the corresponding pipeline of CL HS for surfaces with uniform chromaticity from Chapter 4 (Figure 4.4), there is the addition of the chromaticity calibration procedure feeding into constraint computation at the reconstruction stage, along with the output of photometric
5.2 Surface chromaticity calibration

In this section, a spatio-temporal chromaticity calibration procedure applicable to dynamic as well as static scenes is proposed. The two main stages of the procedure are the initial per-camera chromaticity estimation in the reference frame (Section 5.2.1) and its subsequent temporal propagation (Section 5.2.2) to any new frame provided sufficient overlap with the reference. Both unseen (non-reference) static scenes as well as entire sequences in dynamic scene reconstruction can be served by the spatio-temporal chromaticity calibration approach.

5.2.1 Surface chromaticity estimation

The procedure. In this section, a procedure for per-pixel chromaticity estimation of the reconstructed surface is proposed. Since the three light sources in CL HS are red (R), green (G) and blue (B), the goal of the chromaticity calibration procedure is to compute the triplet \( (p'_{c,R}, p'_{c,G}, p'_{c,B}) \) consisting of three chromaticity coefficients relative to the reference \( (p_{c,R}^{\text{ref}}, p_{c,G}^{\text{ref}}, p_{c,B}^{\text{ref}}) \). The triplet \( (p'_{c,R}, p'_{c,G}, p'_{c,B}) \) describes the relative reflectance behaviour in response to red, green and blue illumination spectra for a visible surface point \( x \) viewed by camera \( C_c \) where \( c = \{1, 2, 3\} \). It is assumed that the illumination spectra relate to the spectral sensor characteristics of the three cameras in the same way.

The calibration method is based on sampling the chromatic response of first a planar object with the chosen reference chromaticity \( (p_{c,R}^{\text{ref}}, p_{c,G}^{\text{ref}}, p_{c,B}^{\text{ref}}) \) and then that of the arbitrarily coloured object to be calibrated. Both the reference and the calibrated objects remain static during sampling to facilitate per-pixel estimation. For sampling of chromatic response, both objects are sequentially exposed to red, green and blue illumination from the same direction.
Chapter 5. CL HS for dynamic scenes with arbitrary spatially-varying chromaticity

a. reference surface sampling  b. calibrated surface sampling  c. per-pixel relative chromaticity estimation

Figure 5.2: Reference and calibrated surface sampling for chromaticity estimation.

by changing colour filters of a single static light source. The colour filters used to sample chromatic response are subsequently intended for data acquisition at reconstruction. The sum of the RGB colour filter spectra defines white illumination in this context. Let us formalise chromaticity estimation per pixel of any \( C_c \) where \( c = \{1, 2, 3\} \) since each pixel in the procedure is calibrated independently.

Consider the planar reference surface in Figure 5.2a being sampled by camera \( C_c \) for any \( c = \{1, 2, 3\} \) in the configuration of CL HS. A surface point on the reference object \( x_{\text{ref}} \) with the orientation \( n_{\text{ref}} \) projects onto the camera sensor pixel defined by \( v^{\text{ref}}_1 \). The point is illuminated by a single light source \( S_l \) in the fixed position whose inherent radiance distribution is sequentially coloured red \((\rho_r)\), green \((\rho_g)\) and blue \((\rho_b)\) using filters for the purpose of chromatic response sampling of \( x_{\text{ref}} \). From Equation 5.1 the image formation equations for the per-channel responses \((i_{c(r),r}^{\text{ref}}, i_{c(g),g}^{\text{ref}}, i_{c(b),b}^{\text{ref}})\) at \( x_{\text{ref}} \) corresponding to the spectrum of illumination in each case are:

\[
\begin{align*}
    i_{c(r),r}^{\text{ref}} &= \rho_r(v_2^{\text{ref}})\sigma_r(v_1^{\text{ref}})p_{c,R}^{\text{ref}}f_d(v_2^{\text{ref}}, v_1^{\text{ref}}) \frac{v_2^{\text{ref}} \cdot n_{\text{ref}}}{|c_2 - x_{\text{ref}}|^2} \\
    i_{c(g),g}^{\text{ref}} &= \rho_g(v_2^{\text{ref}})\sigma_g(v_1^{\text{ref}})p_{c,G}^{\text{ref}}f_d(v_2^{\text{ref}}, v_1^{\text{ref}}) \frac{v_2^{\text{ref}} \cdot n_{\text{ref}}}{|c_2 - x_{\text{ref}}|^2} \\
    i_{c(b),b}^{\text{ref}} &= \rho_b(v_2^{\text{ref}})\sigma_b(v_1^{\text{ref}})p_{c,B}^{\text{ref}}f_d(v_2^{\text{ref}}, v_1^{\text{ref}}) \frac{v_2^{\text{ref}} \cdot n_{\text{ref}}}{|c_2 - x_{\text{ref}}|^2}
\end{align*}
\]  

(5.8)

Note that triplet \((p_{c,R}^{\text{ref}}, p_{c,G}^{\text{ref}}, p_{c,B}^{\text{ref}})\) is the chromaticity of the reference surface. As described in Section 5.1, this reference chromaticity must be the same as the chromaticity of the calibration plane in the photometric calibration procedure from Chapter 4 in order to ensure the reference independence of the CL HS pipeline as a whole. Typically, the reference
tends to be chosen in the white spectrum of colours to maximise the channel response, although theoretically it does not have to be. In addition, the chromaticity of the reference object must be uniform since otherwise there will be per-pixel variations in the reference chromaticity, which would be impossible to reconcile with the reference in the photometric calibration.

Now consider the calibrated surface point \( x \) with orientation \( n \) in Figure 5.2b sampled with the same three coloured radiance distributions \( \rho_r, \rho_g \) and \( \rho_b \) in an identical configuration of camera \( C_c \) and light source \( S_l \). The point-to-sensor projection is defined by vector \( v_1 \) and the direction of illumination by \( v_2 \). Each such point \( x \) on the calibrated object is also characterised by a triplet of image formation equations \( (i_{c(r)}, r, i_{c(g)}, g, i_{c(b)}, b) \) defined by its chromaticity \( (p_{c,R}, p_{c,G}, p_{c,B}) \) in response to the same stimuli:

\[
\begin{align*}
    i_{c(r), r} &= \rho_r(v_2)\sigma_r(v_1)p_{c,R}f_d(v_2, v_1) \frac{v_2 \cdot n}{\|c_2 - x\|^2} \\
    i_{c(g), g} &= \rho_g(v_2)\sigma_g(v_1)p_{c,G}f_d(v_2, v_1) \frac{v_2 \cdot n}{\|c_2 - x\|^2} \\
    i_{c(b), b} &= \rho_b(v_2)\sigma_b(v_1)p_{c,B}f_d(v_2, v_1) \frac{v_2 \cdot n}{\|c_2 - x\|^2}
\end{align*}
\]

Note that \( (p_{c,R}, p_{c,G}, p_{c,B}) \) in Equation 5.9 is an absolute chromaticity triplet independent of any reference.

The camera-to-light source geometry is identical within each triplet of expressions in Equations 5.8 and 5.9 because both the set-up and the scene are static during sampling of each surface. This means that the ratio of any two expressions from Equation 5.8 or 5.9 depends only on the relative \( \rho, \sigma \) products and the corresponding per channel chromaticities. For example, the red-to-green response ratio for the reference surface point \( x_{ref} \) is:

\[
\frac{i_{c(r), r}^{ref}}{i_{c(g), g}^{ref}} = \frac{\rho_r(v_2^{ref})\sigma_r(v_1^{ref})}{\rho_g(v_2^{ref})\sigma_g(v_1^{ref})} \frac{p_{c,R}^{ref}}{p_{c,G}^{ref}}
\]

The equivalent ratio for the calibrated surface point \( x \) is:

\[
\frac{i_{c(r), r}}{i_{c(g), g}} = \frac{\rho_r(v_2)\sigma_r(v_1)p_{c,R}}{\rho_g(v_2)\sigma_g(v_1)p_{c,G}}
\]

Chromaticity is estimated per-pixel in the image domain of \( C_c \). Hence the idea is to link
the reference and the calibrated surface points projecting onto the same pixel as shown in Figure 5.2c. In other words, in the sampling configuration employed the projection vectors are the same: $v^\text{ref}_1 = v_1$, and Equation 5.10 simplifies to:

$$\frac{i^\text{ref}}{i^\text{c,g}} = \frac{\rho_r(v^\text{ref}_2)\sigma_r(v_1)}{\rho_g(v^\text{ref}_2)\sigma_g(v_1)} p^\text{ref}_{c,R} p^\text{ref}_{c,G}$$

(5.12)

meaning that the same point in the sensitivity distributions $\sigma_r$ and $\sigma_g$ applies to the corresponding reference and calibrated surface points, $x^\text{ref}$ and $x$ respectively. However, Figure 5.2c also shows that the illumination vectors $v^\text{ref}_2$ and $v_2$ are clearly not the same because the sampled 3D points $x^\text{ref}$ and $x$ are not identical.

An important simplification can be made to Equations 5.11 and 5.12 given that the coloured radiance distributions $\rho_r$, $\rho_g$ and $\rho_b$ are realised by applying colour filters to a radiance distribution of a single light source $S_l$. Colour filters can be assumed spatially uniform which means that the radiance distributions $\rho_r$, $\rho_g$ and $\rho_b$ differ from each other by a constant scale factor. For example, for one pair of distributions one can write: $\rho_r = k\rho_g$ where $k$ is a constant meaning that the same relationship holds for any two spatially corresponding samples of the distributions. With this in mind, the ratio in Equation 5.12 can be re-written as:

$$\frac{i^\text{ref}}{i^\text{c,g}} = \frac{\rho_g(v^\text{ref}_2)\sigma_r(v_1)}{\rho_g(v^\text{ref}_2)\sigma_g(v_1)} p^\text{ref}_{c,R} p^\text{ref}_{c,G} = k\frac{\sigma_r(v_1)}{\sigma_g(v_1)} p^\text{ref}_{c,R} p^\text{ref}_{c,G}.$$  

(5.13)

Equivalently, Equation 5.11 becomes:

$$\frac{i^\text{c,R}}{i^\text{c,g}} = \frac{k\rho_g(v_2)\sigma_r(v_1)p_{c,R}}{\rho_g(v_2)\sigma_g(v_1)p_{c,G}} = k\frac{\sigma_r(v_1)}{\sigma_g(v_1)} \frac{p_{c,R}}{p_{c,G}}.$$  

(5.14)

As a result of the simplification, Equations 5.13 and 5.14 can be combined by substitution for $k\frac{\sigma_r(v_1)}{\sigma_g(v_1)}$ and one obtains an expression for the ratio of two relative chromaticity components $p^\text{ref}_{c,R}$ and $p^\text{ref}_{c,G}$, defined against the reference chromaticity components $p^\text{ref}_{c,R}$ and $p^\text{ref}_{c,G}$, as a function of directly measurable intensities:

$$\frac{p^\text{c,R}}{p^\text{c,G}} = \frac{i^\text{(c,R)}_r}{i^\text{(c,G)}_g} \frac{i^\text{ref}_{c,R}}{i^\text{ref}_{c,G}}.$$  

(5.15)
5.2. Surface chromaticity calibration

Three such ratios, e.g. \( \frac{p'_{c,R}}{p'_{c,G}} \), \( \frac{p'_{c,G}}{p'_{c,B}} \) and \( \frac{p'_{c,R}}{p'_{c,B}} \):

\[
\begin{align*}
\frac{p'_{c,R}}{p'_{c,G}} &= \frac{i_{c(r),r}^r i_{c(g),g}^g}{i_{c(r),r}^r i_{c(g),g}^g} = c_1 \\
\frac{p'_{c,R}}{p'_{c,B}} &= \frac{i_{c(r),r}^r i_{c(b),b}^b}{i_{c(r),r}^r i_{c(b),b}^b} = c_2 \\
\frac{p'_{c,G}}{p'_{c,B}} &= \frac{i_{c(g),g}^g i_{c(b),b}^b}{i_{c(g),g}^g i_{c(b),b}^b} = c_3
\end{align*}
\] (5.16)

result in three homogeneous constraints constituting a homogeneous linear system of equations:

\[
\begin{pmatrix}
1 & -c_1 & 0 \\
1 & 0 & -c_2 \\
0 & 1 & -c_3
\end{pmatrix}
\begin{pmatrix}
p'_{c,R} \\
p'_{c,G} \\
p'_{c,B}
\end{pmatrix} = 0.
\] (5.17)

The system is solved by SVD decomposition of the constraint matrix with as the solution a normalised vector i.e. the chromaticity coefficient triplet \( [p'_{c,R};p'_{c,G};p'_{c,B}]^T \). As explained in Section 5.1, chromaticity describes only the relative inter-channel relationship, not the absolute intensities, which means that multiple colours map onto the same chromaticity (e.g. all greyscale values are the same in terms of chromaticity). The estimation procedure describes each set of colours with the same inter-channel relationship by a single colour from the set, specifically the one corresponding to the normalised vector \( [p'_{c,R};p'_{c,G};p'_{c,B}]^T \). For example, due to this intensity ambiguity, all greyscale colours map onto triplet \( [\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}] \), which is the normalised vector expressing inter-channel equality. The representing colour can be changed to any member of the colour set, for example the maximum intensity colour (all greyscale values would map to white \( [1,1,1] \) in this case).

It should be stressed that disambiguation of colours with the same chromaticity is irrelevant for reconstruction by CL HS. The normal constraint of CL HS in Equation 5.2 is homogeneous meaning that any consistent scaling of chromaticity coefficients cancels out. Provided the inter-channel relationship is preserved, the reconstructed point can, without any consequence to the reconstruction constraint, in fact measure any channel intensity triplet within the colour family of the given inter-channel relationship. The important issue is that both cameras involved in the constraint sample the same inter-channel relationship.
The reference object response effectively calibrates for the photometric properties of the set-up cancelling the dependency in the calibrated surface expression in Equation 5.14 to recover equipment independent estimates. Any observed discrepancy in the chromaticity estimate of corresponding points between cameras \( C_c \) for \( c = \{1, 2, 3\} \) can have two explanations. Firstly, in theory there can be a difference in the way the cameras are matched to the colour filters spectrally which scales spatial sensitivity distributions in a wavelength-dependent way, possibly inconsistent between the reference and the calibrated surfaces. However, in practice since in this work three identical cameras are used, it is highly unlikely there would be such spectral sensitivity differences between them. Secondly, the intensities sampled on the calibrated surface of certain optical properties can be corrupted by sub-surface scattering phenomena (see Section 5.3) in which case the ratio of reference responses in Equation 5.12 no longer accurately describes the relationship between chromaticity and intensity ratios in Equation 5.14 via the shared equipment characteristics leading to camera dependence.

The inter-channel relationship expressed by the relative chromaticity can be converted to an RGB colour for visualisation of the estimated chromatic properties of the surface. The set of RGB values for the visible to \( C_c \) points (i.e. per image pixel) of the calibrated surface constitutes a chromaticity map (for examples see Figure 5.9). The chromaticity map should not be confused with a colour map: there will not be an exact match in appearance of the chromaticity map to the intensity image as due to the brightness ambiguity the actual colour cannot be recovered. Moreover, the appearance of the chromaticity map will depend on the rule adopted as to which colour family member is to represent the family’s inter-channel relationship (e.g. the maximum brightness colour rule). The rule should of course be kept consistent for all pixels in the map.

For surfaces with uniform chromaticity, the per-pixel chromaticity map can be averaged within the silhouette mask of the reconstructed object. The average is computed for each channel individually. The advantage of estimating chromaticity as an average for a surface \( a \ priori \) known to be uniform is the elimination of minor spatial variations of per-pixel estimates for potentially smoother reconstructions. The limitation of averaging is obviously the fact that the approach only lends itself to chromaticity estimation of surfaces with uniform chromaticity. Note that the restriction is not as limiting as assuming the same
5.2. Surface chromaticity calibration

colour: the colours are allowed to vary arbitrarily within the family of colours with a
given inter-channel relationship e.g. any greyscale pattern on a white surface would be
indistinguishable from white. When the inter-channel chromatic relationship varies across
the surface the per-pixel chromaticity map must be used instead.

5.2.2 Surface chromaticity propagation

The chromaticity estimation procedure described in the previous section is static, with
a non-instantaneous acquisition, mapping chromaticity triplets to pixel locations in one
particular reference frame. For reconstruction of dynamic sequences with spatially varying
chromaticity per-pixel colour information for every frame is needed. In this section, a
chromaticity propagation procedure is proposed to infer chromaticity in each frame of the
dynamic sequence from the original reference frame chromaticity map.

Let $I_s$ denote the reference frame which is the source of chromaticity information. For
a given camera $C_c$, the reference frame is characterised by a chromaticity map $M_{\text{chrom}}^{\text{ref}}$
where there is an estimate of the chromaticity triplet for each pixel within the silhouette
of the object of interest. The dynamic sequence to be reconstructed consists of a set of $N$
frames $[I_{d,1}, I_{d,2}, ..., I_{d,N}]$ featuring the same object of interest. The procedure of surface
chromaticity propagation consists in deriving from the reference chromaticity map $M_{\text{chrom}}^{\text{ref}}$
the set of $N$ chromaticity maps $[M_{\text{chrom},1}^d, M_{\text{chrom},2}^d, ..., M_{\text{chrom},N}^d]$ corresponding to the
dynamic sequence frames.

**Alignment.** As pre-processing, $M_{\text{chrom}}^{\text{ref}}$ must be aligned to the object pose in $I_{d,1}$, the
first frame of the dynamic sequence, in order to form the starting point for chromaticity
propagation. As $M_{\text{chrom}}^{\text{ref}}$ is per-pixel of $I_s$, it is equivalent to align $I_s$ to $I_{d,1}$ instead.
Subsequently, by applying the same alignment transformation to $M_{\text{chrom}}^{\text{ref}}$, the chromaticity
map of the first frame in the dynamic sequence $M_{\text{chrom},1}^d$ is obtained.

In the simplest case, chromaticity estimation immediately precedes dynamic capture and
$I_s$ will be roughly the same as $I_{d,1}$. Unaided perfect alignment however occurs only in the
case of static reconstruction of $I_s$, the frame used for chromaticity estimation. Any minor
misalignment of pose can be corrected using optical flow techniques.

However, one would not wish to be limited to reconstruction of just the tailored dynamic
sequence and hence re-use of chromaticity data for dynamic sequences featuring the same
Figure 5.3: Alignment of the reference frame $I_s$ to the first dynamic sequence frame $I_{d,1}$ for a given camera. The superimposed feature grids were manually initialised in the source and target frames. Local (piecewise) homography with bi-cubic interpolation is used to warp the source frame onto the target frame.

object in different initial poses is desirable. The problem of object pose variation with such untailored dynamic sequences is solved by warping $I_s$ (the source) to $I_{d,1}$ (the target) provided a sufficient degree of overlap between the two. The warping procedure involves initialisation of corresponding feature points in both the source and the target and a transformation to align the features. The nature of the transform used depends on the surface being aligned. For (near-)planar surfaces, a global homography may be sufficient whereas surfaces with local curvatures require more sophisticated forms of warping i.e. a local (piecewise) homography. For alignment of non-planar surfaces, the user is required to define a coarse grid whose vertices correspond to scene features. Warping is then performed by a piecewise homography involving bi-cubic interpolation. Non-planar alignment is illustrated in Figure 5.3 that shows the source and target intensity images with the manually initialised feature grids superimposed and the resultant warping of the source onto the target. The approach has been found capable of coping with a visually substantial difference between the source and target. The alignment stage of the propagation process is the only part of the pipeline requiring manual interaction for feature matching. This requires minimal user effort and in return provides the ability to process multiple dynamic sequences with substantially varying initial poses from a single chromaticity map estimate $M_{ref}^{chrom}$.

**Propagation.** The alignment procedure described in the previous section generates the chromaticity map $M_{chrom,1}^{d}$ of the first dynamic sequence frame $I_{d,1}$ from the reference chromaticity map $M_{ref}^{chrom}$. The main idea of the subsequent propagation procedure is to compute chromaticity for all the subsequent frames of the dynamic sequence by propagating $M_{chrom,1}^{d}$ throughout. The propagation from $I_{d,1}$ onwards is achieved by sequentially computing optical flow between adjacent pairs of intensity frames ($I_{d,n}, I_{d,n+1}$) and us-
5.2. Surface chromaticity calibration

Figure 5.4: Chromaticity propagation between a pair of adjacent frames. The backward and forward flow maps are computed based on intensity images. The chromaticity of frame $I_{n+1}$ is derived from that of frame $I_n$ using the backward flow map.

Dense optical flow tracking for chromaticity propagation is performed on the intensity images $[I_{d,1}, I_{d,2}, ..., I_{d,N}]$ of the dynamic sequence for each camera separately. Specifically, the efficient GPU implementation of optical flow from [119] is used, which has the large displacement optical flow (LDOF) [120] in its core. LDOF is a variational technique with a continuous energy functional whose optimisation is embedded into a coarse-to-fine framework allowing one to estimate large displacements even for the smaller scene components [120]. The ability of the algorithm to cover a wide range of displacement amplitudes permits freedom in the choice of motion speed and frame rates of the tracked dynamic scenes.

Tracking produces per-pixel flow maps [121] that are used in this work to propagate the chromaticity map estimate $M_{\text{chrom},1}^d$ aligned to the first frame $I_{d,1}$ throughout the dynamic sequence, effectively establishing a mapping from the calibrated chromaticity $M_{\text{chrom}}^{ref}$ to each frame. To ensure spatial coverage completeness of the frame-to-frame mapping, backward flows are utilised rather than forward flows, meaning that each pixel of $M_{d,n+1}^d$ of the current frame $I_{n+1}$ is assigned the chromaticity triplet (if defined) of the quantised to the nearest pixel back-projection to $M_{d,n}^d$ of the previous frame $I_n$. The nearest neighbour approach is chosen over higher order interpolation in order to avoid colour blurring at region boundaries of the chromaticity map. To illustrate the process, Figure 5.4 shows, for one particular camera, 1. two adjacent intensity frames with a substantial relative motion;
2. the propagation of chromaticity between the two and 3. the corresponding backward and forward flow maps illustrated as a flow direction heatmap.

The process of propagation hence provides the sought spatially-varying chromaticity estimates per-frame of a dynamic sequence.

5.3 Error analysis

The chromaticity estimation procedure devised for CL HS is tailored to the technique and therefore logically makes the same assumptions about the radiometric properties of the surface. Specifically, HS relies on the Bidirectional Reflectance Distribution Function (BRDF) being an accurate abstraction of surface reflectance. In reality, the BRDF is the simplest model for the phenomena at the interface between media whereby the light ray incident at a surface point is reflected from that same point remaining on the incidence side of the interface. The Bidirectional Transmission Distribution Function (BTDF), complimentary to the BRDF, describes light scattering upon exitance, on the other side of the surface. Together the BRDF and BTDF form the Bidirectional Scattering Distribution Function (BSDF) which is said to be a fair model of light scattering behaviour for: 1. very thin surfaces with both BRDF and BTDF (see Figure 5.5a.); 2. certain types of clear materials (glass, acrylic etc.) with primarily BTDF and 3. certain types of opaque materials with primarily BRDF. While the BRDF-only surfaces are the target group of HS, completely clear materials are out of its scope. For an infinitesimally thin surface, in the limit, the points of incidence and exitance are identical for both BRDF and BTDF in the BSDF model. Such thin surfaces photometrically described by the BSDF present a challenge for HS discussed in more detail later on in this section.

For thicker objects also the BSDF is inadequate because sub-surface scattering [122] may start contributing to exiting illumination as well. Whether sub-surface scattering occurs depends on the surface material type. Sub-surface scattering (see Figure 5.5b.) complicates the model because the point of light incidence is no longer the same as the point of exitance (which is the case for both BRDF and BTDF in the limit when the surface is one point thick). The model taking sub-surface scattering into account is known as the Bidirectional Surface Scattering Reflectance Distribution Function (BSSRDF) [123]. By separating the points of incidence and exitance, the BSSRDF (Figure 5.5b.) essentially increases the com-
5.3. Error analysis

The terms sub-surface scattering and translucency (also shown in Figure 5.5b.) are often used interchangeably to describe light taking random paths within a surface of finite thickness before re-emerging on either side of it. In this work the choice is made to view translucency primarily as a superset of transparency whereby, unlike transparency, light rays transported through the surface do not necessarily follow Snell’s laws of refraction. Hence, also unlike transparency, translucency does not permit sharp image formation behind the surface since the light rays passing through are scattered instead of bundled along the path defined by Snell’s laws. Sub-surface scattering on the other hand will be referred to more in the context of sub-surface reflection to the side of incidence.

Hence the two main sources of error causing the radiometric behaviour to deviate from the BRDF model are transparency and sub-surface scattering/translucency (Figures 5.5a. and 5.5b. respectively). These phenomena alter the intensity measurements in a way not described by the BRDF which upsets both the HS constraint in Equation 5.5 by breaking Helmholtz reciprocity and the chromaticity estimation ratio in Equation 5.15. Ultimately, the surfaces showing a substantial amount of these phenomena are not in the
Chapter 5. CL HS for dynamic scenes with arbitrary spatially-varying chromaticity

scope of CL HS. However, it will be shown in this chapter that some level of transparency/translucency (e.g. a thin cloth) and sub-surface scattering (e.g. human skin) is tolerated by the proposed algorithm which still produces plausible reconstructions. The remainder of this section illustrates how the BRDF assumption is violated by transparency and sub-surface scattering/translucency and discusses the consequences of the violation for the Helmholtz reciprocity constraint and the chromaticity estimation procedure.

5.3.1 Transparency

The BTDF models how light passes through the surface without internal scattering to emerge on the other side, i.e. at the opposite media interface, with a certain distribution as a function of viewing angle (Figure 5.5a). At the incidence interface, upon entering the surface, the light ray of frequencies not coinciding with the resonant frequencies of the material will be refracted instead of absorbed, which constitutes material transparency [122]. Refraction depends on the angle of incidence and the respective refraction indices of the media. For the growing angle of incidence, the reflected ray is gradually becoming stronger while the refracted one is becoming dimmer. Beyond the critical angle of incidence the ray undergoes total internal reflection without any transmission. Due to this directional dependence, light transmission introduced by the property of transparency results in variable amounts of measured illumination loss depending on the location of the light source. The variable illumination loss means that the maximum reflection response attainable at a surface point becomes dependent on the angle of incidence. Furthermore, the property of variable illumination loss by transmission (which can be lumped together with absorption losses) as a function of the angle of incidence breaks the fundamental reciprocity property of the BRDF: its directional components \( f_d(v_2, v_1) \) and \( f_d(v_1, v_2) \) at a surface point will no longer be equal since, except for the special case when \( v_2 \) and \( v_1 \) form the same angle of incidence to the surface, the illumination rays will undergo different absorption processes. The resultant inconsistency between \( i_{1,2} \) and \( i_{2,1} \) due to different transmission by refraction is uncalibrated for by the BRDF and hence perturbs the normal constraint vector.

For thick surfaces translucency discussed further on will dominate and pure transparency is characteristic only of special kinds of materials such as glass, polyethylene etc. that are beyond the scope of CL HS. For thin surfaces (e.g. fabrics, paper) of arbitrary material
there will be a transparency component to sub-surface transport along with a translucency component to account for scattering in the material layer of finite thickness. Exactly how dominant the two components are relative to one another would depend on the thickness and type of material. The corruption of chromaticity estimate will be modelled in the next section for the more generic case of translucency.

5.3.2 Sub-surface scattering/translucency

The presence of sub-surface scattering (Figure 5.5b) is an example of complex optical behaviour exhibited by some surfaces whereby instead of simply reflecting off the surface, the light ray penetrates the material, bounces randomly within its structure and re-emerges at an arbitrary point. Sub-surface scattering described by the BSSRDF results in translucency which is a more generic form of transparency. Unlike transparency described previously, translucency does not enforce Snell’s laws and light upon crossing the interface re-emerges after scattering on either side of the surface of arbitrary thickness. In this section the implications of sub-surface scattering for the chromaticity estimate are considered.

Consider a surface point \( x_1 \) observed by a camera at the viewing angle defined by vector \( v_1 \) and illuminated at the angle of incidence defined by vector \( -v_2 \) (Figure 5.5b), the reverse of the vector from the surface point to the light source. Traditionally, reflectance behaviour is described by the BRDF \( f_r(v_2, v_1) \) that only models the phenomena at the media boundary treating any illumination penetrating the object as irrevocably lost by absorption. The BSSRDF is a more generic abstraction for reflectance behaviour superseding the BRDF. The BSSRDF models sub-surface scattering explicitly as a contributor to the observed intensity whereas the BRDF at best adjusts the balance between reflection and absorption depending on whether there is a net gain or loss of energy by sub-surface scattering. The BSSRDF is not characterised by Helmholtz reciprocity as the BRDF and hence breaks the normal constraint of HS as has been discussed for the special case of transparency.

So, for the previously defined surface point \( x_1 \), the reflection of the incidence light ray \( -v_2 \) at the surface point is modelled by \( \sum_{\forall x_1'} f'(x_1, v_2, x_1', v_1') \) where each \( f'(x_1, v_2, x_1', v_1') \) corresponds to a single scatter component re-emerging from within the surface at point \( x_1' \) in the corresponding direction \( v_1' \) to the camera. The important special case component with \( x_1 = x_1' \) approximates direct surface reflectance. Due to the gradual energy loss, the
Chapter 5. CL HS for dynamic scenes with arbitrary spatially-varying chromaticity

effect of sub-surface scattering from the ray incident at \( \mathbf{x}_1 \) on exitant illumination will be localised to a patch centred at \( \mathbf{x}_1 \) with some radius \( ||\mathbf{x}_1 - \mathbf{x}'_1||_{\text{max}} \) (Figure 5.5b). Function \( f'(\mathbf{x}_1, \mathbf{v}_2, \mathbf{x}'_1, \mathbf{v}'_1) \) describing a single scatter component is the BSSRDF i.e. a function of incident illumination entering at \( \mathbf{x}_1 \) at the angle defined by \( \mathbf{v}_2 \) and exitant at some \( \mathbf{x}'_1 \) along \( \mathbf{v}'_1 \) towards the camera.

The image formation equations (Equation 5.9) can now be modified to take sub-surface scattering into account. In the formulation of channel intensities in Equation 5.18, the image-pixel-defining vector \( \mathbf{v}_1 \) at point \( \mathbf{x}_1 \) is kept constant while summing the contributions to the observed exitant illumination, via direct surface reflectance and sub-surface scattering, of incident rays \( \mathbf{v}'_2 \) from set \( \mathcal{V}_2 \) entering at corresponding surface points \( \mathbf{x}'_2 \) within the patch radius \( ||\mathbf{x}_1 - \mathbf{x}'_2||_{\text{max}} \) from \( \mathbf{x}_1 \):

\[
\begin{align*}
    i_{c(r)} &= \sum_{\mathbf{v}'_2 \in \mathcal{V}_2} \rho_r(\mathbf{v}'_2) \sigma_r(\mathbf{v}_1) f'(\mathbf{x}'_2, \mathbf{v}'_2, \mathbf{x}_1, \mathbf{v}_1, \omega_r) \frac{\mathbf{v}'_2 \cdot \mathbf{n}}{||\mathbf{c}_2 - \mathbf{x}'_2||^2} \\
    i_{c(g)} &= \sum_{\mathbf{v}'_2 \in \mathcal{V}_2} \rho_g(\mathbf{v}'_2) \sigma_g(\mathbf{v}_1) f'(\mathbf{x}'_2, \mathbf{v}'_2, \mathbf{x}_1, \mathbf{v}_1, \omega_g) \frac{\mathbf{v}'_2 \cdot \mathbf{n}}{||\mathbf{c}_2 - \mathbf{x}'_2||^2} \\
    i_{c(b)} &= \sum_{\mathbf{v}'_2 \in \mathcal{V}_2} \rho_b(\mathbf{v}'_2) \sigma_b(\mathbf{v}_1) f'(\mathbf{x}'_2, \mathbf{v}'_2, \mathbf{x}_1, \mathbf{v}_1, \omega_b) \frac{\mathbf{v}'_2 \cdot \mathbf{n}}{||\mathbf{c}_2 - \mathbf{x}'_2||^2}
\end{align*}
\]

The intensity triplet \((i_{c(r)}, i_{c(g)}, i_{c(b)})\) corresponds to surface point \( \mathbf{x}_1 \). Since in this case sampling is performed with multi-chromatic illumination, the incident wavelength \( \omega \) dependence of the BSSRDF must be made explicit: \( f'(\mathbf{x}'_2, \mathbf{v}'_2, \mathbf{x}_1, \mathbf{v}_1, \omega) \). The radiance-sample-identifying vector \( \mathbf{v}'_2 \) varies depending on the point of incidence. The sensitivity sample is fixed by \( \mathbf{v}_1 \) in the expressions as a single point of exitance \( \mathbf{x}_1 \) is being observed.
Let us now make the special case term corresponding to \( x_2' = x_1 \) explicit in the sum:

\[
\begin{align*}
    i_{c(r),r} & = \rho_r(v_2)\sigma_r(v_1)f'(x_1, v_2, x_1, x_1, \omega_r) \frac{v_2 \cdot n}{||c_2 - x_1||^2} + \\
    & + \sum_{v_2' \in V : v_2' \neq x_1} \rho_r(v_2')\sigma_r(v_1)f'(x_2', v_2', x_1, x_1, \omega_r) \frac{v_2' \cdot n}{||c_2 - x_1||^2} \\
    i_{c(g),g} & = \rho_g(v_2)\sigma_g(v_1)f'(x_1, v_2, x_1, x_1, \omega_g) \frac{v_2 \cdot n}{||c_2 - x_1||^2} + \\
    & + \sum_{v_2' \in V : v_2' \neq x_1} \rho_g(v_2')\sigma_g(v_1)f'(x_2', v_2', x_1, x_1, \omega_g) \frac{v_2' \cdot n}{||c_2 - x_1||^2} \\
    i_{c(b),b} & = \rho_b(v_2)\sigma_b(v_1)f'(x_1, v_2, x_1, x_1, \omega_b) \frac{v_2 \cdot n}{||c_2 - x_1||^2} + \\
    & + \sum_{v_2' \in V : v_2' \neq x_1} \rho_b(v_2')\sigma_b(v_1)f'(x_2', v_2', x_1, x_1, \omega_b) \frac{v_2' \cdot n}{||c_2 - x_1||^2}
\end{align*}
\]

(5.19)

Let us assume that, like the BRDF before, the BSSRDF with \( x_2' = x_1 \) can be split into a wavelength-dependent component and a directional component: \( f'(x_1, v_2, x_1, x_1, \omega) = f'(\omega)f_d(x_1, v_2, x_1, v_1) \). The directional component implicitly models as absorption the total loss of illumination by sub-surface scattering at observed point \( x_1 \) as a function of sampling geometry. The wavelength-dependent component of the BSSRDF \( f'(\omega) \) can be modelled as the product \( \gamma_l(v_2)\rho_{c,l} \). In the decomposition \( \rho_{c,l} \) is the topical chromaticity coefficient of surface point \( x_1 \) when viewed by camera \( c = \{1, 2, 3\} \) and illuminated by light source \( l = \{R, G, B\} \). Further the illumination-direction-dependent scaling parameter \( \gamma_l(v_2) \) accounts for the differences in the illumination loss by absorption through sub-surface scattering between channels, modelling the relative absorptive scattering response to illumination of different spectra \( l = \{R, G, B\} \), spectrally variable due to the sub-surface matter’s chromatic and other physical properties (e.g. resonant frequency etc.). In the chromaticity estimation procedure, the illumination direction \( v_2 \) at the sampled point is constant between channels in Equation 5.19 and between cameras. Hence, parameter \( \gamma_l(v_2) \) in this case is wavelength dependent only: \( \gamma_l(v_2) = \gamma_l \). When \( x_2' = x_1 \) as in \( f_d(x_1, v_2, x_1, v_1) \), i.e. the points of incidence and exitance are identical, direct surface reflectance at \( x_1 \) can be assumed to dominate over the contribution of the single sub-surface scattering component as the result of illumination entering and exiting through the same point \( x_1 \) as, geometrically speaking, the size of point \( x_1 \) tends to zero in the limit. The additive contribution to intensity by
subsurface scattering in Equation 5.19 is considered to arise from the incidence points other than \(x_1\) in the radius patch \(||x_1 - x'_2||_{\text{max}}\) and only the intensity distortion by absorption is modelled here by \(f'(x_1, v_2, x_1, v_1, \omega)\) where \(x'_2 = x_1\).

Substituting the above definitions into Equation 5.19 and re-arranging gives:

\[
i_{c(r),r} = \sum_{v'_2 \in V_2, x'_2 \neq x_1} \rho_r(v'_2)\sigma_r(v_1)f'(x'_2, v'_2, x_1, v_1, \omega_r)\frac{v'_2 \cdot n}{||c_2 - x'_2||^2} = \\
= \rho_r(v_2)\sigma_r(v_1)\gamma_{RPc,R} f'_d(x_1, v_2, x_1, v_1) \frac{v'_2 \cdot n}{||c_2 - x'_2||^2}
\]

\[
i_{c(g),g} = \sum_{v'_2 \in V_2, x'_2 \neq x_1} \rho_g(v'_2)\sigma_g(v_1)f'(x'_2, v'_2, x_1, v_1, \omega_g)\frac{v'_2 \cdot n}{||c_2 - x'_2||^2} = \\
= \rho_g(v_2)\sigma_g(v_1)\gamma_{GPc,G} f'_d(x_1, v_2, x_1, v_1) \frac{v'_2 \cdot n}{||c_2 - x'_2||^2}
\]

\[
i_{c(b),b} = \sum_{v'_2 \in V_2, x'_2 \neq x_1} \rho_b(v'_2)\sigma_b(v_1)f'(x'_2, v'_2, x_1, v_1, \omega_b)\frac{v'_2 \cdot n}{||c_2 - x'_2||^2} = \\
= \rho_b(v_2)\sigma_b(v_1)\gamma_{BPc,B} f'_d(x_1, v_2, x_1, v_1) \frac{v'_2 \cdot n}{||c_2 - x'_2||^2}
\]

Taking a ratio of channel intensity expressions in Equation 5.20, exactly like in Equation 5.11, one obtains:

\[
\frac{i_{c(r),r} - \sum_{v'_2 \in V_2, x'_2 \neq x_1} \rho_r(v'_2)\sigma_r(v_1)f'(x'_2, v'_2, x_1, v_1, \omega_r)\frac{v'_2 \cdot n}{||c_2 - x'_2||^2}}{i_{c(g),g} - \sum_{v'_2 \in V_2, x'_2 \neq x_1} \rho_g(v'_2)\sigma_g(v_1)f'(x'_2, v'_2, x_1, v_1, \omega_g)\frac{v'_2 \cdot n}{||c_2 - x'_2||^2}} = \frac{\rho_r(v_2)\sigma_r(v_1)\gamma_{RPc,R}}{\rho_g(v_2)\sigma_g(v_1)\gamma_{GPc,G}} = k \frac{\sigma_r(v_1)\gamma_{RPc,R}}{\sigma_g(v_1)\gamma_{GPc,G}}
\]

as the geometry of sampling, as well as the directional component of the BSSRDF \(f'_d(x_1, v_2, x_1, v_1)\), cancel out being identical between channel expressions and the radiance distributions are uniformly scaled relative to one another. Substituting for \(k \frac{\sigma_r(v_1)}{\sigma_g(v_1)}\) using Equation 5.12, exactly as in Equation 5.15, results in a constraint relating two relative per-channel chromaticity coefficients with compensation for sub-surface scattering via
the unperturbed intensity ratio \( \left( \frac{i_{c(r),r}}{i_{c(g),g}} \right) \): \[
\frac{p'_{c,R}}{p'_{c,G}} = \frac{\frac{p_{c,R}}{p_{c,G}}}{\frac{p_{c,R}}{p_{c,G}}} = \left( \frac{i_{c(r),r}}{i_{c(g),g}} \right) \frac{\frac{\sigma_{f}(v_1)}{\sigma_{g}(v_1)}}{\frac{\sigma_{f}(v_1)}{\sigma_{g}(v_1)}} \right) \approx \left( \frac{i_{c(r),r}}{i_{c(g),g}} \right) \frac{\sigma_{f}(v_1)}{\sigma_{g}(v_1)} \right) \right) \]

5.3. Error analysis

Several important comments should be made. Firstly, the perturbation of the intensity ratio by sub-surface scattering at surface point \( x_1 \) can theoretically be compensated for knowing 1. the contribution to the measured reflectance on each channel individually of all the relevant sub-surface scattering terms from other points of incidence (i.e. \( x_2 \neq x_1 \)) and 2. the ratio \( \frac{\sigma_{f}}{\sigma_{g}} \) modelling per-channel inconsistencies in absorption by sub-surface scattering at point \( x_1 \). Comparing Equations 5.15 and 5.22 it is clear that the directly observed intensity ratio \( \frac{i_{c(r),r}}{i_{c(g),g}} \) used in the former equation is a perturbed version of \( \left( \frac{i_{c(r),r}}{i_{c(g),g}} \right) \). In practice, it is impossible to observe the unperturbed ratio \( \left( \frac{i_{c(r),r}}{i_{c(g),g}} \right) \) directly, hence the assumption of negligible sub-surface scattering is made approximating \( \left( \frac{i_{c(r),r}}{i_{c(g),g}} \right) \) by the perturbed ratio \( \frac{i_{c(r),r}}{i_{c(g),g}} \). Note from Equation 5.22 that this approximation requires that there are no contributions to the reflectance at \( x_1 \) from the other points of incidence and that absorption is either completely unaffected by sub-surface scattering or affected in the same way for both channels. The analysis for simplicity also assumes that the increased absorption by sub-surface scattering can be modelled as a function of sampling geometry and wavelength and there is no random statistical element in it. If a surface violates these assumptions, the approximation \( \frac{i_{c(r),r}}{i_{c(g),g}} \approx \left( \frac{i_{c(r),r}}{i_{c(g),g}} \right) \) is inaccurate. In this case, \( \frac{i_{c(r),r}}{i_{c(g),g}} \) retains the dependency on the photometric properties of the acquisition configuration via the unaccounted sum scattering terms. The scattering terms are unaccounted for as the ratio of reference intensities with no sub-surface scattering describing \( k \sigma_{f}(v_1)/\sigma_{g}(v_1) \) only eliminates camera dependence if the surface intensities \( i_{c,R, c,G} \) are also a result of direct surface reflectance only.

When the acquisition equipment properties no longer cancel out by sampling the reference response, the estimated chromaticites will be to some extent relative to the comparative dominance, via the light source radiance and camera sensitivity, of the three channels in the set-up. Even more importantly, the sum of sub-surface scattering term itself differs per
camera because of the directional dependence of the BSSRDF \( f'(x_2', v_2', x_1, v_1, \omega) \). As a result, if the object exhibits substantial sub-surface scattering, the chromaticity estimate of a point becomes camera-dependent and the chromaticity maps estimated will appear inconsistent. Since the \( \gamma_R \) to \( \gamma_G \) ratio is purely wavelength dependent, it cannot be the cause of chromaticity estimate discrepancy between the cameras as the same spectra are consistently used for sampling. Also note that the absorption by sub-surface scattering at \( x_1 \) as a function of sampling geometry per camera does not play a role in any such discrepancy either as the directional components \( f_d'(x_1, v_2, x_1, v_1) \) cancel out in the ratio of two channel intensities.

Note that the directional dependence of the BSSRDF discussed in this section is not characterised by Helmholtz reciprocity like that of the BRDF, which violates the core premise of Helmholtz Stereopsis as a reconstruction method in addition to introducing the inconsistencies in the chromaticity estimates.

### 5.3.3 Reference chromaticity

Let us briefly highlight a few properties of the reference surface, used in the chromaticity calibration procedure as shown in Figure 5.2, that are important for accuracy of the entire pipeline. The choice of reference chromaticity, as well as colour, is arbitrary although there is a logical preference for maximising the response on all channels. On the other hand, sensor saturation in the reference measurement is a source of error: all cameras of the set-up should be adjusted accordingly to limit the occurrence of saturated patches in the image characterised by uncalibrated photometric behaviour of equipment. Further, the chromaticity must be as uniform as possible across the reference surface to avoid per-pixel reference variations unreconcilable between the chromaticity and photometric calibration procedures. Recall that the chromaticity estimates of the object to be reconstructed are relative to the reference chromaticity (Equation 5.15) which cancels out in Equation 5.5 against the same dependency of the photometric calibration. Planarity of the reference surface is preferred in the chromaticity estimation to avoid self-shadowing through any non-planar features negatively impacting chromatic uniformity of the reference. Except for self-shadowing prevention, there is no reason for planarity of the reference surface in the chromaticity estimation procedure. Finally, reflectance of the reference surface is accurately
5.4. Implementation

5.4.1 The pipeline and reference chromaticity

Colour Helmholtz Stereopsis (CL HS) extended to scenes with arbitrary spatially-varying chromaticity can be schematically summarised by the pipeline in Figure 5.6. The crucial addition compared to the CL HS for uniform chromaticity scenes in Figure 4.9 is the spatio-temporal chromaticity calibration procedure described in detail in Section 5.2 and illustrated in Figure 5.7. The reference chromaticity must be consistent between the photometric and chromaticity calibration procedures. In the photometric calibration the reference chromaticity is a characteristic of the target plane with markers. In the chromaticity calibration it is the reference relative to which chromatic response is measured. The photometric calibration fully describes the reflectance response of objects with reference chromaticity. Given the reference consistency, for chromatically different objects it does so also when scaled by the estimated relative (to the reference) chromaticity. Without the reference consistency, an extra conversion reconciling the different references of photometric and chromaticity calibration procedures would be required. To ensure reference consistency, the same surface material is used in both procedures, which in theory can be any colour, not necessarily white. In practice, it is not advisable to choose reference chromaticities with too low a response on any of the channels as it introduces instability into the linear system in Equation 5.17. Such low response also means that measurements are taken with a low signal-to-noise ratio. In this work a white reference is used for simplicity although any chromaticity with a sufficiently high, but not necessarily balanced, response on all channels is suitable. Further, the calibration plane is matte to avoid sensor saturations that would interfere with the calibration of the spatial parameter distributions.

5.4.2 Cross-talk

The so-called cross-talk occurs when a fraction of a signal intended for one channel is received on another channel. In the presented system based on wavelength-multiplexing cross-talk is the RGB signals exciting the wrong camera sensors. The energy may be thus described by the BRDF model (i.e. no sub-surface transport).
Chapter 5. CL HS for dynamic scenes with arbitrary spatially-varying chromaticity

Figure 5.6: Coarse-to-fine CL HS pipeline for reconstruction of complex dynamic scenes with arbitrarily spatially-varying chromaticity.

Figure 5.7: Spatio-temporal chromaticity calibration pipeline. Chromaticity in the reference frame is estimated per camera. The resultant chromaticity maps (i.e., spatial chromaticity distributions) are aligned with the reconstructed view. If a dynamic sequence is being reconstructed, the aligned chromaticity maps are temporally propagated throughout the dynamic sequence using dense point propagation.
lost to other channel sensors or received from an unintended stimulus. Either way the received signal on each individual channel can be distorted by cross-talk.

A way to spatially estimate cross-talk is to measure the response of the three multi-spectral Helmholtz cameras in the reconstruction configuration to red, green and blue light reflected from a white surface. Assuming a perfectly white surface, the signal of a given colour in theory should excite a response only in the corresponding channel. The sum of responses on the other two channels is cross-talk. Measuring the response in the reconstruction configuration allows one to estimate a spatial cross-talk error distribution (i.e. a cross-talk map) for a real-life scenario. Such a map does not allow to correct for cross-talk at reconstruction as the chromatic properties of the reconstructed surface will change the error
Table 5.1: Cross-talk statistics: rms and maximum cross-talk percentages of the spatial distributions presented in Figure 5.8.

<table>
<thead>
<tr>
<th></th>
<th>HS camera 1 (red)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>stimulus</td>
<td>rms cross-talk [%]</td>
<td>max. cross-talk [%]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>green</td>
<td>2.14</td>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>blue</td>
<td>0.87</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HS camera 2 (green)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stimulus</td>
<td>rms cross-talk [%]</td>
<td>max. cross-talk [%]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>red</td>
<td>1.54</td>
<td>53.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>blue</td>
<td>0.88</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HS camera 3 (blue)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stimulus</td>
<td>rms cross-talk [%]</td>
<td>max. cross-talk [%]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>red</td>
<td>1.34</td>
<td>53.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>green</td>
<td>2.12</td>
<td>60</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

distribution. However, a cross-talk map can serve as a good indication of the quality of the Helmholtz camera configuration as will be demonstrated.

Let us consider an example reconstruction configuration of three multi-spectral Helmholtz cameras that helps provide practical suggestions for a better set-up in terms of cross-talk. For each camera the cross-talk is measured under two signal wavelengths (the third wavelength is the one the Helmholtz camera emits itself and hence is irrelevant). Cross-talk is represented as a spatial distribution of the percentage leaked signal from the total signal strength. Figure 5.8 shows the six distributions in the three camera reconstruction configuration whereas the corresponding statistical metrics of the distributions are given in Table 5.1. Generally speaking, the rms cross-talk observed in this configuration is about 1 – 2% (which is at most 15 intensity levels but in practice closer to 5 intensity levels as the total signal strength is unlikely to exceed 255 by much with a single coloured stimulus). However, the actual effect the percentage error has on the reconstruction depends on the overall signal strength as well. As the general trend cross-talk percentage tends to increase towards the outskirts of the frame as the total signal strength there tends to be substantially weaker with essentially a low signal-to-(cross-talk) noise ratio. The drastic cross-talk maxima given in Table 5.1 occur at near-zero signal intensity levels. To minimise the effect of cross-talk in the reconstruction the signal-to-noise ratio must be kept high in the region of interest. In other words the scene should be optimally illuminated. From the distribution presented in Figure 5.8 it is clear that the blue light is oriented much more advantageously with the scene
well-lit in a clearly defined spotlight where the cross-talk is under 0.5%. The orientations of the green and red light sources are less advantageous.

With a proper light source positioning resembling that of the blue light in the example configuration, cross-talk is unlikely to be a major source of error for the system. In setting up the reconstruction configuration care must be taken to ensure that the scene receives the maximum amount of light from all the three light sources in order to minimise percentage cross-talk (or equivalently to maximise the signal-to-noise ratio) in the reconstruction volume. Such a configuration can be easily achieved during scene framing by collocating the spatial centre of the reconstruction volume with the optical spotlight axes of all the projectors. In addition to the illumination configuration, the RGB sensitivity spectra and the patterns of their mutual overlap characteristic of the specific cameras used are a factor influencing observed cross-talk percentages as well.

5.5 Evaluation

The validation of the methodology presented in this chapter consists of several stages. Firstly, in Section 5.5.1 chromaticity estimation is assessed for a range of objects with varying radiometric properties. Static scene reconstruction with arbitrary spatially-varying chromaticity in Section 5.5.2 is incorporated to test the pipeline directly with the per-pixel chromaticity estimation without any tracking errors that may be introduced in the chromaticity propagation procedure. The two datasets included in the section are a near-white plaster statue of a monster head with local spatial discolouration (“Monster”) and a multi-coloured toy-dog made of primarily plastic and cloth (“Slinky”). The extensive dynamic scene reconstruction section (Section 5.5.3) is split into several sections testing the effect of the entire CL HS pipeline on scenes with: 1. uniform (near-)reference chromaticity similar to those in Chapter 4 (“Tea towel” - white tea towel cloth deformation); 2. uniform non-reference chromaticity (“Jumper”- knitted fabric deformation); 3. non-reference pseudo-uniform\(^3\) chromaticity (“Hand” - hand gestures) and, finally, the ultimate case of 4. arbitrary spatially-varying chromaticity (“Face” - facial expressions). In conclusion of the validation, Section 5.5.4 shows the performance of the proposed algorithm on two pho-

\(^3\)a property whereby the spatial material-level average still approximates chromaticity adequately despite its noticeable non-uniformity.
tometrically challenging datasets with extreme sub-surface scattering and sensor saturation characteristics. These datasets break the fundamental assumptions of 3D reconstruction in general, not just those of the proposed algorithm, and define the current limitations on photometric complexity.

5.5.1 Chromaticity estimation

In this section, the estimated per-camera spatial chromaticity distributions for this chapter’s datasets are presented and discussed in order to demonstrate the wide applicability scope of the chromaticity estimation algorithm. Firstly, one goal is to verify the accuracy of surface chromaticity capture regardless of the frequency of its spatial variation. Secondly, the procedure’s independence of the BRDF is to be shown. Finally, in support of the error discussion in Section 5.3, the effect of transparency/translucency and sub-surface scattering in the BSSRDF on the chromaticity estimates is illustrated. The BSSRDF of Monster and (most of) Slinky is characterised by direct surface reflectance with negligible sub-surface scattering and translucency. In other words, the BSSRDF of these surfaces reduces to the classical BRDF modelling topical surface reflectance only. For these two objects the algorithm generates three per-camera chromaticity maps that are consistent in appearance for corresponding object parts.

Chromaticity describes the inter-channel relationship not their absolute intensities. Hence, one can speak of colour families with the same inter-channel relationship where all members map onto the same chromaticity triplet. In Figure 5.9 each per-pixel chromaticity triplet is visualised by the highest intensity colour in the family obtained by dividing the triplet by its largest element. This type of visualisation means that all greyscale colours (i.e. the normalised chromaticity vector $\left[\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right]$ or equal intensities on all channels) will be represented as white $[255, 255, 255]$, rather then grey at $[147, 147, 147]$.

The dominant chromaticity of Monster is represented by a white colour consistent between the cameras. The per-camera chromatic setting variations, clearly exposed in the corresponding intensity images, have no effect on the consistency of the chromaticity estimates because the photometric camera properties are accounted for using the channel-to-channel response on the reference object (Equation 5.13). Chromaticity in the shadowed regions of

---

2the transparent circular compartment in the hind leg with a mirror on the back-wall is out of scope.
5.5. Evaluation

Figure 5.9: Spatial chromaticity distribution of test scenes for cameras C₁, C₂, and C₃. Intensity images under unfiltered projector illuminations are shown for reference of appearance. For visualisation, each per-pixel chromaticity triplet is represented by the maximum intensity colour in the family mapping to it (which corresponds to per-pixel scaling by the reciprocal of the largest coefficient in the triplet).
the surface (the mouth, the outer curvatures of the head etc.) cannot be reliably estimated as the signal is sampled at extremely low intensities, with a low signal-to-noise ratio. The light source used for sampling in the chromaticity calibration procedure is collocated with camera $C_1$. This explains why shadows affect mainly the intensity images of $C_2$ and $C_3$. The estimates in the shadowed regions reveal the colour biases in the settings of the black level of the cameras. The observed bias when sampling the (near-) black level is based on inter-channel intensity measurements of low magnitudes but, due to the relative normalised nature of the chromaticity estimate, the dominance of one channel (e.g. red on $C_2$) is still represented by the maximum brightness colour with the sampled inter-channel relationship.

Like Monster, Slinky is also characterised by a high degree of consistency in the chromaticity estimates between the cameras due to the absence of sub-surface reflectance behaviour. The visualisation colours in the chromaticity maps do not exactly match the intensity images being an arbitrary choice of the inter-channel relationship representative but it is evident that per-region colour families are accurately computed. Once again, the lack of channel balance in the camera black level is apparent when estimating chromaticity in a shadow. For example, the red chromaticity, to which the cameras are clearly biased, bleeds into the purple and green regions on Slinky’s back and tail. However, apart from these shadowed regions, the algorithm can be said to successfully separate regions of the spatially-varying chromaticity distribution.

In contrast to Monster and Slinky, there is a substantial degree of variation in the chromaticity maps of Hand and Face between the cameras. The assumption of predominantly topical surface reflectance and the reduction of the BSSRDF to the BRDF is not valid for human skin as there is a known existence of sub-surface scattering for this surface type. In Section 5.3, the hypothetical ratio of two channel intensities unperturbed by sub-surface scattering $\left(\frac{i_{c_1}(l_1)}{i_{c_2}(l_2)}\right)$ was modelled in terms of the directly measurable intensities on such a surface: $\left(\frac{i_{c_1}(l_1)}{i_{c_2}(l_2)}\right) = \frac{\gamma_{c_2}}{\gamma_{c_1}} \left(\frac{\sum v'_2 \in V_2, v'_2 \neq v_1, \sigma_{c_1} f'_{v_2}(v_1, v_1, \omega_1)}{\sum v'_2 \in V_2, v'_2 \neq v_1, \sigma_{c_2} f'_{v_2}(v_1, v_1, \omega_1)}\right)$. The chromaticity estimates are not consistent between cameras for surfaces with sub-surface scattering because the perturbed intensity ratio $\frac{i_{c_1}(l_1)}{i_{c_2}(l_2)}$ measured directly is used in the procedure instead of the unavailable unperturbed ratio $\left(\frac{i_{c_1}(l_1)}{i_{c_2}(l_2)}\right)$. In the model showing the relationship between the two, there is the important summation term representing
reflectance by sub-surface scattering from other incidence points. Such a term is contained in both the directly measured (perturbed) channel intensities, $i_{c(l_1),l_1}$ and $i_{c(l_2),l_2}$. The terms are dependent on the sampling geometry, as well as on the wavelength due to the heterogeneous sub-surface point chromaticities encountered during scattering, via $f'(x'_2, v'_2, x_1, v_1, \omega)$ which means they will not be the same for the different cameras sampling from different positions. In addition, the terms retain the dependence on the photometric properties of the acquisition equipment ($\sigma_l$ and $\rho_l$), which in the absence of sub-surface scattering are eliminated from the equation using the reference response. This uncalibrated dependence would make the chromaticity estimates camera-dependent even if the additive contribution to reflectance by sub-surface scattering were consistent for the three cameras.

Finally, Towel and Jumper, being relatively thin textiles, suffer from reflectance complexity due to transparency and/or translucency. Both the global differences in the average chromaticity estimate between the cameras and the unexpected spatial variations within each estimated chromaticity map (see the maps of cameras $C_2$ and $C_3$ especially) are observed. The phenomenon of pure transparency cannot be the cause of these discrepancies since it only models the variable illumination loss by absorption as a function of the angle of incidence and the wavelength. Since all cameras sample the response at a given point with the illumination coming from the same direction, the loss will be consistent. For a given camera, there can be a difference in the absorption losses due to transparency between different surface points. However, since the channel intensities per point experience a consistent absorption loss by transparency as a function of incidence angle, the chromaticity estimate consistency between surface points will be unaffected (unless the reflectance is reduced to negligible values, which will not happen for the type of material considered). Finally, although refraction governing transparency is wavelength dependent, any perturbation in the absorption losses due to sampling with different frequencies will be consistent between cameras and surface points as well and thus not the cause of neither global nor local estimate discrepancies observed. Despite being unaffected by pure transparency, the consistency of the chromaticity estimate for these datasets locally and between the cameras is corrupted with the additive contribution to the intensity by sub-surface scattering observed on the incidence side due to the property of translucency which both textiles are
The Towel, Jumper, Hand and Face datasets all to some extent violate the assumption of pure BRDF without sub-surface reflectance phenomena made by the chromaticity estimation algorithm. The violation shows as the discussed spatial chromaticity estimate errors and inconsistencies between the cameras. Since the BSSRDF is not reciprocal like the BRDF the normal constraint of CL HS will not be strictly valid for these surfaces, which is also bound to introduce errors. However, in practice, as will be shown in the following sections, some sub-surface reflectance behaviour appears not to be prohibitive of plausible reconstruction by CL HS. Also the reconstructions of objects strictly abiding by the BRDF assumption will be presented.

5.5.2 Static scenes

The purpose of evaluation on static scenes with arbitrary spatially varying chromaticity is to assess the first stage of the chromaticity calibration procedure i.e. surface chromaticity estimation (Section 5.2.1) independently of the subsequent surface chromaticity propagation stage (Section 5.2.2) essential for dynamic scene reconstruction. The two static test objects used are Monster and Slinky. Monster is made of plaster having a Lambertian reflectance model but with its chromaticity generally off-white, not matching the chromatic reference used, and non-uniform due to staining. Slinky is made of a variety of complex materials such as non-Lambertian plastic, signal-scattering “furry” cloth, transparent plastic and a mirror. In addition the toy is characterised by highly variable chromaticity with some pure colours and by complex geometries such as the “high-frequency” middle section of the body. The intensity images with the corresponding estimated chromaticity maps per camera for the two datasets are presented in Figure 5.9. Figure 5.10 subsequently shows the visual hull and the reconstructions for the fully calibrated CtF ML HS and CtF MRF HS (photometric calibration (of Helmholtz cameras) (PhotoCalib) and chromaticity (surface) calibration (ChromCalib)), the partially calibrated CtF MRF HS (i.e. PhotoCalib only) as well as the completely uncalibrated CtF MRF HS.

Qualitatively, at the first glance in Figure 5.10, Monster reconstructions for all three CtF MRF HS pipelines look very similar. However, the heat maps in Figure 5.11 reveal substantial structural mesh differences manifesting themselves in both depth and normal
5.5. Evaluation

<table>
<thead>
<tr>
<th>Visual hull</th>
<th>CtF ML HS with PhotoCalib and ChromCalib</th>
<th>CtF MRF HS without calibration</th>
<th>CtF MRF HS with PhotoCalib</th>
<th>CtF MRF HS with PhotoCalib and ChromCalib</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Visual hull" /></td>
<td><img src="image2" alt="CtF ML HS" /></td>
<td><img src="image3" alt="CtF MRF HS" /></td>
<td><img src="image4" alt="CtF MRF HS with PhotoCalib" /></td>
<td><img src="image5" alt="CtF MRF HS with PhotoCalib and ChromCalib" /></td>
</tr>
<tr>
<td><img src="image6" alt="Flat shading" /></td>
<td><img src="image7" alt="a." /></td>
<td><img src="image8" alt="b." /></td>
<td><img src="image9" alt="a." /></td>
<td><img src="image10" alt="b." /></td>
</tr>
<tr>
<td><img src="image11" alt="Smooth shading" /></td>
<td><img src="image12" alt="a." /></td>
<td><img src="image13" alt="b." /></td>
<td><img src="image14" alt="a." /></td>
<td><img src="image15" alt="b." /></td>
</tr>
</tbody>
</table>

Figure 5.10: Reconstruction of static scenes with arbitrary spatially-varying chromaticity. Reconstruction cores compared: Shape-from-silhouette (visual hull), CtF ML HS with PhotoCalib and ChromCalib (fully calibrated), uncalibrated CtF MRF HS, CtF MRF HS with PhotoCalib only and fully calibrated CtF MRF HS. Initial sampling resolution: 1 mm/0.5 mm (spatially/depthwise) for Monster and 1 mm/1 mm for Slinky. Integration for CtF MRF HS: NoInt; Integration for CtF ML HS: a. NoInt; b. Poisson Surface Reconstruction.
estimates. Visually, the heat maps expose epicentres of local depth deviation of up to 1 cm (e.g. the lip) from the fully calibrated result. Despite appearances, the rms depth deviation of the completely uncalibrated result (2.1 mm) is not substantially different from the score for the partially (only photometrically) calibrated one (1.9 mm). The corresponding rms normal deviations of 15.8° and 17.6° respectively even suggest that the totally uncalibrated result is closer to the fully calibrated than the partially calibrated one. The important point in this context is the fact of incompleteness of photometric calibration alone for objects with chromaticity unlike the reference. Partial calibration cannot be expected to facilitate improvement in performance of the pipeline and in fact coincidently may result in its degradation (or, given the lack of ground truth, more deviation from the fully calibrated result). The only reason why some decrease in the depth deviation from the fully calibrated result is observed (an improvement trend not supported by the normals however) with the introduction of photometric calibration in the case of Monster is the local similarity of the object’s chromaticity to the reference.

The necessity of full calibration is obvious for Slinky with its clearly more pronounced spatial chromaticity variations. As with Monster, the reconstruction improvements facilitated by partial calibration alone are minor as the photometric and chromaticity calibration procedures are indivisible. Quantitatively, both partially and completely uncalibrated results deviate from the fully calibrated reconstruction by the rms value of 1.5 cm in depth and a drastic rms normal error of 54°. This extreme normal distortion is evident in the Slinky reconstructions with smooth shading in Figure 5.10 as well: the global average orientation of the normals is clearly skewed resulting in a darker general shading.

Bayesian CL HS with or without calibration clearly outperforms its fully calibrated maximum likelihood variant. The NoInt method of arranging computed vertices into a mesh is not suitable for CtF ML HS whose depth maps are extremely noisy with the input of three reciprocal pairs. While showing the results of CtF ML HS with NoInt for consistency, the corresponding meshes obtained by Poisson Surface Reconstruction (PR) are also shown to give an idea of the quality given a more suitable integration method. Also with PR as the integration method, CtF ML HS reconstructions are distorted to the extreme. The results support the conclusion stated earlier in Chapter 4 that the standard maximum likelihood Helmholtz Stereopsis is not suitable in the context of dynamic scene reconstruction with
### 5.5. Evaluation

<table>
<thead>
<tr>
<th>depth error</th>
<th>normal error</th>
</tr>
</thead>
<tbody>
<tr>
<td>uncalibrated CtF MRF HS</td>
<td>uncalibrated CtF MRF HS</td>
</tr>
<tr>
<td>CtF MRF HS with PhotoCalib</td>
<td>CtF MRF HS with PhotoCalib</td>
</tr>
</tbody>
</table>

Figure 5.11: Heat maps showing spatial distribution of the depth and normal deviation (in mm and ° respectively), of the under-calibrated CtF MRF HS results from the fully calibrated CtF MRF HS.

<table>
<thead>
<tr>
<th>depth error</th>
<th>normal error</th>
</tr>
</thead>
<tbody>
<tr>
<td>uncalibrated CtF MRF HS</td>
<td>uncalibrated CtF MRF HS</td>
</tr>
<tr>
<td>CtF MRF HS with PhotoCalib</td>
<td>CtF MRF HS with PhotoCalib</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>depth error</th>
<th>normal error</th>
</tr>
</thead>
<tbody>
<tr>
<td>uncalibrated CtF MRF HS</td>
<td>uncalibrated CtF MRF HS</td>
</tr>
<tr>
<td>CtF MRF HS with PhotoCalib</td>
<td>CtF MRF HS with PhotoCalib</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>depth error</th>
<th>normal error</th>
</tr>
</thead>
<tbody>
<tr>
<td>uncalibrated CtF MRF HS</td>
<td>uncalibrated CtF MRF HS</td>
</tr>
<tr>
<td>CtF MRF HS with PhotoCalib</td>
<td>CtF MRF HS with PhotoCalib</td>
</tr>
</tbody>
</table>
Chapter 5. CL HS for dynamic scenes with arbitrary spatially-varying chromaticity

wavelength-multiplexing.

On the other hand, the proposed CtF MRF HS shows both the global shape accuracy and an impressive resolution of structural detail facilitated by the coarse-to-fine approach. For example, notice the resolution of the wrinkles and boils as well as the scratched up plaster on the right-hand side of Monster. For Slinky the method copes with the specular plastic, the signal scattering “furry” structure of the face and the high frequency geometry in the spring-like torso. Even the small balls in the hind-leg with the mirror behind are resolved. Slinky remains a challenging object with layers of geometric and photometric complexity and there is room for improvement in the reconstruction accuracy. However, given the complexity and only a three-image input, the performance on Slinky is pleasing. Together the two presented static datasets provide strong evidence for the claim of CL HS with the core of CtF MRF HS being a competitive method for accurate multi-chromatic object reconstruction.

5.5.3 Dynamic scenes

Having validated CL HS with the core of CtF MRF HS on multi-chromatic static scenes, the evaluation can now focus on the target application of dynamic scenes with arbitrary reflectance properties. Unlike static scenes, dynamic scenes call for temporal propagation of static per-pixel chromaticity estimate, possibly with a pre-processing stage of alignment of the chromaticity calibration frame to the first frame of the dynamic sequence if the two are not the same. Hence, dynamic scenes add an extra level of complexity evaluating the pipeline with the full spatio-temporal chromaticity calibration procedure (Figure 5.7).

The evaluation for dynamic scenes consists of four stages. Firstly, as a bridge to Chapter 4, the effect of chromaticity calibration on the reconstruction of scenes with uniform near-reference chromaticity is studied. Such scenes were reconstructed in Chapter 4 without chromaticity calibration. The goal is to verify whether fine-tuning the chromaticity via the spatio-temporal calibration procedure improves the result without introducing any instability. For this stage the chosen dataset is a white tea towel (“Tea Towel”) with a fine surface structure of raised squares being substantially globally deformed in the dynamic sequence.

The second stage of the dynamic scene evaluation features an object with uniform chro-
maticity substantially different from the reference. For this purpose the fabric of a woollen
jumper ("Jumper") is deformed in the second sequence. The example is interesting because
the uniformity of chromaticity in this case is approximate due to the natural variations in
the colour of wool as well as the periodic gaps in the knitwear introducing black dot noise.
As has been discussed previously, the fabric of Jumper, as well as Tea Towel, exhibits
translucency violating the assumption of Helmholtz reciprocity. In this evaluation, Jumper
is treated as a scene with uniform non-reference chromaticity.

Thirdly, the assumption of chromaticity uniformity is further challenged by reconstructing
individual frames from a sequence featuring hand gestures ("Hand"). The frames are re-
constructed with the uniform (non-reference) chromaticity assumption instead of per-pixel
calibration because the chosen poses differ substantially from the chromaticity calibration
frame with new uncalibrated for parts being exposed. As chromaticity cannot be propa-
gated in this case due to the missing information in the calibration shot as well as due to the
tracking difficulties given the extreme pose variation, the reconstruction results obtained
using chromaticity calibration by the spatial average are presented.

Finally, the ultimate case of arbitrary spatially-varying chromaticity is considered. These
scenes require per-pixel chromaticity estimation and propagation throughout the sequence.
The dataset used for this stage is a dynamic sequence of continuously changing facial ex-
pressions ("Face") including a large variety of deformations performed at different speeds.
"Face" is photometrically a challenging dataset due to the highly spatially varying chro-
maticity (no make-up was used to even out complexion), its non-Lambertian reflectance
model, sub-surface scattering and shadowing by facial features such as the nose. The high
variability of poses and motion speeds introduce temporal complexity to the sequence.

**Uniform arbitrary chromaticity**

The three datasets in this category, i.e. Tea Towel, Jumper and Hand, are chromatically
characterised by the spatial average of per-pixel chromaticity map in the region of interest.
For Tea Towel the stage of calibration-to-reconstruction alignment is skipped as redundant
due to the material’s highly uniform chromaticity throughout. To obtain the chromaticity
map in this case, the entire manually defined reconstruction silhouette is simply filled with
the mean chromaticity of the spatial per-pixel distribution in the calibration shot. Despite
Chapter 5. CL HS for dynamic scenes with arbitrary spatially-varying chromaticity

the fact that Hand is not characterised by chromaticity uniformity in the same way as Tea Towel, the same method is used for this dataset because the chromaticity estimate alignment/propagation to the region of interest is not possible with the challenging novel hand poses considered. For Jumper, only piecewise uniform in chromaticity, the region of interest on the fabric is initialised by aligning the calibration frame to the reconstructed frame using global homography based on 20 points. Local homography is not needed in this case because the fabric is approximately planar in both the source and the target poses of the transformation.

The workflow for obtaining such a patch-aligned material-level chromaticity estimate in the region of interest is illustrated in Figure 5.12 on the Jumper dataset. Per-pixel chromaticity is estimated in the reference (calibration) frame (note that the chromaticity triplet is normalised to unity in Figure 5.12 and not scaled by the maximum RGB component as in Figure 5.9). The homography also defines the region of interest for the reconstruction which can be accurately chromatically described using the average from the calibrated material patch. So, the same transformation is performed on the chromaticity map where per-pixel estimates are uniformly replaced by the single average. In case there is the knowledge of material uniformity throughout, the region of interest is irrelevant and full-frame reconstruction can be performed. In this case, the choice has been made to restrict reconstruction to the region of interest because there is some non-uniformity of material observed later on in the sequence.

In order to propagate the region of interest throughout the sequence the per-pixel intensity tracking results by optical flow are utilised. So, in case of material-level chromaticity estimation, the region of interest tracking is a tool for establishing per-frame silhouettes, propagating the region of interest in which the average chromaticity is known to be valid without actually updating the average in any way. For Tea Towel without the region of interest initialisation and with the chromaticity assumed to be valid everywhere on the fabric, the fabric silhouettes for the subsequent frames can be selected manually for one-off frame reconstructions. For dynamic sequences, the original manual silhouette initialisation can be propagated based on intensity tracking or, alternatively, reconstruction can be confined to the deformation within a fixed reconstruction volume.

Let us now look at the reconstruction results for Tea Towel, Jumper and Hand for a set of
5.5. Evaluation

<table>
<thead>
<tr>
<th>$I_s$</th>
<th>$I_s$ warped onto $I_{d,1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>camera 1</td>
<td>camera 2</td>
</tr>
</tbody>
</table>

**Intensity**

<table>
<thead>
<tr>
<th>per-pixel $M_{\text{chrom}}^{ref}$ in $I_s$</th>
<th>$M_{\text{chrom}}^{d,1}$ averaged $M_{\text{chrom}}^{ref}$ warped onto $I_{d,1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>camera 1</td>
<td>camera 2</td>
</tr>
</tbody>
</table>

Chromaticity

Figure 5.12: Material-level chromaticity estimation with alignment. Left column: Intensity images for cameras $C_1$, $C_2$ and $C_3$ under unfiltered projector light and the corresponding per-pixel chromaticity maps $M_{\text{chrom}}^{ref}$ for the reference (calibration) frame $I_s$. Right column: intensity alignment of $I_s$ to $I_{d,1}$, the first frame of the dynamic sequence, initialising the region of interest for reconstruction and the (in this case) average-based chromaticity map $M_{\text{chrom}}^{d,1}$ per camera in the region of interest.

frame 26 | frame 41 | frame 89 | frame 144 | frame 185

Figure 5.13: Intensity images of $C_1$ (top), $C_2$ (middle) and $C_3$ (bottom) for some representative frames from the dynamic Tea Towel sequence with arbitrary uniform chromaticity.

representative frames from the dynamic sequences.

**Tea Towel - uniform (near-)reference chromaticity.**

Similarly to the datasets presented in the previous chapter, Tea Towel is of uniform chromaticity closely approximating the reference. From the flat-shaded reconstructions in Figure 5.14 best showcasing pure geometry the improvement with the introduction of photometric calibration is evident. The reconstructions by the uncalibrated pipeline are retracted, showing a clearly erroneous surface orientation both globally and of some local features. For this dataset there is no evident reconstruction improvement by adding explicit chromaticity calibration to the pipeline.
Chapter 5. CL HS for dynamic scenes with arbitrary spatially-varying chromaticity

Figure 5.14: Reconstruction of complex dynamic scenes with arbitrary uniform chromaticity using CL HS: Tea Towel. Reconstruction cores compared: fully calibrated CtF ML HS, uncalibrated CtF MRF HS, partially calibrated CtF MRF HS (PhotoCalib only), and fully calibrated CtF MRF HS with average-based (material-level) chromaticity (proposed). Initial sampling resolution: 3 mm/0.5 mm; Final sampling resolution: 0.375 mm/0.03125 mm. Integration: NoInt shown for all with flat shading (additionally Poisson surface reconstruction added for ML HS only).
Figure 5.15: The reconstruction results from Figure 5.14 shown in smooth shading (taking photometric normals in account).
On detailed mesh comparison the fully calibrated and partially calibrated results are similar but not identical. However, both looking plausible, it is not obvious which geometry is closer to reality. Chromaticity calibration is not essential for datasets like Tea Towel because the chromaticity characteristics closely approach the reference. However, since in practice one cannot assess visually to full precision how similar the reconstructed object’s chromaticity is to the reference, there is no harm in fine-tuning with an explicit chromaticity estimate provided it can be estimated accurately.

The general reconstruction quality of this dataset is exceptionally good. The elaborate global deformations unconstrained by any form of rigidity are accurately captured and the fine details of the surface structure are clearly visible already in the flat shading (Figure 5.14) and indeed very well resolved in the smooth shading (Figure 5.15). The accuracy is in a marked contrast to the failing performance of the standard fully calibrated CtF ML HS with both NoInt and Poisson Surface Reconstruction for the final mesh assembly. The conclusions drawn here based on the presented reconstructions of individual frames from the dynamic sequence are supported by the reconstruction of the full sequence with the resolution of the final iteration at 0.375 mm/0.0635 mm (spatially/depthwise). The reconstructed sequence captures fabric deformation and vertical displacement with temporal consistency.

**Jumper - uniform non-reference chromaticity.** Five representative frames from the dynamic Jumper sequence are shown in Figure 5.16. These frames include both in-place deformation and lateral translation of the fabric. The corresponding reconstructions for the same set of reconstruction cores are shown in flat and smooth shading respectively in Figures 5.17 and 5.18. Flat shading showcases the true geometry. The level of fabric geometry reconstructed by CtF MRF HS is very detailed. Even in the flat shading individual wool threads can be discerned. The corresponding reconstructions rendered in the smooth shading almost look textured in their appearance. In contrast, in the smooth-shaded CtF ML HS, one can barely make out the global shape of the jumper patches, regardless of the method of surface integration employed.

Comparing the CtF MRF HS results with different levels of calibration closely a clear pattern is evident. The partially calibrated CtF MRF HS (photometric calibration only) tends to produce reconstructions with inflated features. In fact, the completely uncalibrated CtF MRF HS produces visually more accurate reconstructions than the partially
section 5.5. Evaluation

Figure 5.16: Intensity images of $C_1$ (top), $C_2$ (middle) and $C_3$ (bottom) for some representative frames from the dynamic Jumper sequence with arbitrary uniform chromaticity.

calibrated one. As observed in the evaluation on static data, the photometric and chromaticity calibration procedures are inseparable parts of a single whole with the end goal of the reference chromaticity cancellation in Equation 5.5 with the resultant remainder of the absolute scene chromaticity and photometric parameter distributions of the acquisition equipment. The procedure can also be seen as a customisation of the default reference photometric calibration to the chromatic properties of the scene. The only exception to this inseparability of photometric and chromaticity calibration is the case of scenes with near-reference chromaticity (see Tea Towel) where the reference photometric calibration is already representative of the scene without any adjustment.

Without chromaticity calibration in the Jumper sequence photometric calibration introduces abnormal feature-inflating instability. Chromaticity calibration definitely moderates the feature inflation of the partially calibrated result settling in-between the uncalibrated and partially calibrated results. Without ground truth it is not obvious whether the fully calibrated or the uncalibrated reconstructed sequence is more accurate as both look plausible. The accuracy of the uncalibrated sequence depends on the validity of the assumption that the photometric characteristics of the Helmholtz cameras are near-identical in the region of interest and the chromaticity is largely characterised by an equal response on all channels. The accuracy of the fully calibrated result depends on how representative the average chromaticity is of the entire surface of the fabric for the overall calibration balance with the relative photometric calibration. Figure 5.9, showing the per-pixel chromaticity maps for each viewpoint, exposes non-uniformities in the chromaticity distribution. However,
Chapter 5. CL HS for dynamic scenes with arbitrary spatially-varying chromaticity

Figure 5.17: Reconstruction of complex dynamic scenes with arbitrary uniform chromaticity using CL HS: Jumper. Reconstructions cores compared: fully calibrated CtF ML HS, uncalibrated CtF MRF HS, partially calibrated CtF MRF HS (PhotoCalib only) and fully calibrated CtF MRF HS with average-based (material-level) chromaticity (proposed). Initial sampling resolution: 3 mm/0.00125 mm. Final sampling resolution: 0.375 mm/0.0003125 mm. Integration: NoInt shown for all with flat shading (additionally Poisson Surface Reconstruction added for CtF ML HS only).
Figure 5.18: The reconstruction results from Figure 5.17 shown in smooth shading (taking photometric normals in account).
these are brought about by the knitwear translucency as well as the transparencies through the gaps in the woven fabric and should best be ignored. Since these non-uniformities contribute to the average it is but an approximation of the true uniform colour of Jumper. Nonetheless, the assumption of the fully calibrated reconstruction pipeline is more likely, instilling more confidence in the accuracy of its result.

As part of evaluation, a 200-frame dynamic Jumper deformation sequence has been reconstructed with the proposed CtF MRF HS at all calibration levels. Note that the frames in the dynamic sequence are reconstructed with the final (i.e. last iteration of the coarse-to-fine framework) resolution of 0.75 mm/0.125 mm (spatially/depthwise) instead of 0.375 mm/0.03125 mm of the individually reconstructed frames in Figures 5.17 and 5.18. Temporally, all three CtF MRF HS reconstructions are equally stable showing a realistically smooth object deformation. Per-frame global accuracy on close examination in the sequence shows the same trend as the individual frame results discussed previously in terms of the uncalibrated results having the flattest features and the partially calibrated the most inflated ones with the fully calibrated result somewhere in-between the two.

**Hand - non-reference pseudo-uniform chromaticity.** Figure 5.19 shows five novel poses distinctly different from the pose adopted at calibration (see the corresponding per-pixel chromaticity maps in Figure 5.9). The chromaticity of skin is not strictly uniform but, unlike Face in the next section with its chromatically distinct features, Hand reconstruction can still be attempted using the average material-level chromaticity. The use of an average is inevitable in this case given the novel poses in Figure 5.19 untrackable from the calibration pose. By computing such an average, any chromaticity estimation artefacts are mitigated. Some such artefacts (e.g. the magenta thumb in the viewpoint of $C_1$) are caused by an insufficient dynamic intensity range of the camera with, as a consequence, the inability to accurately capture the channels with low response (blue and green in this case) leading to the observed RGB imbalance. The cameras are not affected equally as each viewpoint observes its own illumination reflection scenario. Further differences between chromaticity maps of individual viewpoints are caused by sub-surface scattering that is also directional in nature (see Section 5.3).

The reconstructions obtained with the fully calibrated CtF MRF HS using material-level average chromaticity estimates are presented in flat and smooth shading in Figures 5.20
5.5. Evaluation

Figure 5.19: Intensity images of $C_1$ (top), $C_2$ (middle) and $C_3$ (bottom) for some representative frames from the dynamic Hand sequence with non-reference chromaticity approaching non-uniformity.

and 5.21 respectively. The pure geometry without photometric normals in Figure 5.20 reveals a high degree of detail with pronounced skin wrinkles as well as the global shape accuracy. The last two frames suffer from self-occlusions in the 3-camera configuration. However, as a 2.5D reconstruction within the scope, the reconstructions are globally accurate. Under-calibrated pipelines produce distorted geometries: note, for instance, the grotesquely exaggerated indentation and the flesh fold in frames 616 or 654 respectively. Another representative distortion is in frame 650: there is a distinct dent at the tip of the nail in the under-calibrated results whereas in the mesh by the proposed fully calibrated pipeline the nail plate is substantially straighter. The under-calibrated geometries are also
Chapter 5. CL HS for dynamic scenes with arbitrary spatially-varying chromaticity

Figure 5.20: Reconstruction of complex scenes with non-reference pseudo-uniform chromaticity using CL HS: Hand. Reconstruction cores compared: fully calibrated CtF MRF HS (proposed); Initial sampling resolution: 1mm/0.5mm; Final sampling resolution: 0.25mm/0.03125mm; Integration: NoInt shown for all with flat shading.

substantially less smooth. The results validate the need for calibration whereas the perceived deterioration of the artefacts by partial calibration (e.g. frames 615 and 654) once again supports the claim of the indivisibility of the photometric and chromaticity calibration procedures.

The smooth shaded meshes reveal further details of the skin surface bas-relief (e.g. frame 963) and the underlying anatomy (e.g. frame 615). The typical global distortion of the photometric normal orientation is apparent in the under-calibrated results along with the loss of detail resolution.

Several key achievements of the algorithm on this dataset should be pointed out. Firstly, note the overall non-Lambertian glossiness of skin itself successfully processed by the reconstruction algorithm. Secondly, finger nail reconstruction with CtF MRF HS is quite remarkable considering the total lack of chromatic nail calibration, the nail’s specular reflectance model and its tendency to translucency. Note that the quality of nail reconstruction by the under-calibrated CtF MRF HS in smooth shading is deceptive as the normals are accidentally plausible while the global surface orientation at the tip is hugely distorted.
5.5. Evaluation

In this case the pure geometries in Figure 5.20 are truer indicators of performance. In summary, Hand reconstructions are good considering that the scenes have not been chromatically calibrated per-pixel.

**Spatially-varying chromaticity**

Figure 5.9 shows the appearance of the face in the calibration reference frame of each camera and the corresponding estimated per-pixel chromaticity maps essential for accurate reconstruction given spatially varying chromaticity. Each chromaticity map captures spatial skin colour variation at the scale ranging from facial features to local hyper-pigmentation. As discussed previously, any per-camera inconsistencies in the chromaticity maps can be attributed to sub-surface scattering phenomena of the surface.

**Temporal propagation of chromaticity.** The calibration pose in Figure 5.9 is significantly different from the pose in the first frame of the dynamic sequence (see Figure 5.22). The result of aligning the reference chromaticity maps to the first dynamic sequence frame by warping as described in Section 5.2.2 is the leftmost set of images in Figure 5.23. The figure also shows the optical-flow-based propagation of the aligned chromaticity maps through
Chapter 5. CL HS for dynamic scenes with arbitrary spatially-varying chromaticity

Figure 5.22: Intensity images of $C_1$ (top), $C_2$ (middle) and $C_3$ (bottom) for some representative frames from the dynamic Face sequence with arbitrary spatially varying chromaticity.

The sequence. Optical flow tracking shows stability with the chromaticity maps staying constant in the initial static part of the sequence. At the same time during the part of rapid motion from frame 29, the chromaticity maps evolve accordingly with the chromaticity estimates propagated to the new shots. The interior of the mouth unseen in the reference (calibration) frame acquires the colour of the chromatically similar lips, which is not a bad guess for the unavailable information.

Figure 5.23: Temporal chromaticity propagation within the dynamic Face sequence for $C_1$ (top), $C_2$ (middle) and $C_3$ (bottom).

When propagation is continued further into the sequence, tracking will eventually accumulate drift manifesting itself by the chromatic region boundaries corresponding to facial features going fuzzy. Drift, which is essentially the result of numerical error accumula-
5.5. Evaluation

tion, is a known problem of tracking over long sequences. One obvious simple remedy for the problem is re-initialising chromaticity by performing the alignment of the reference chromaticity map to several frames at different temporal points in the sequence. Such a re-initialisation is however only possible if there are enough frames substantially similar to the reference throughout the sequence. Although at the cost of increased computational complexity, another idea to alleviate the problem could be the non-sequential tracking approaches [125], [126]. The basic idea is to break up the sequence into groups of similar poses arranged in a tree structure. Tracking is then performed along this tree where the frames are arranged by similarity rather than sequentially.

Reconstruction. As for the datasets presented previously, Figure 5.24 compares reconstruction results using the fully calibrated CtF ML HS and the proposed CtF MRF HS at different levels of calibration. Figure 5.25 presents the same results with smooth shading to allow comparison of the photometric normal accuracy in addition to the geometry best showcased by flat shading of Figure 5.24.

As in previously presented results, fully calibrated ML HS performs poorly at most giving a vague idea of the global shape. Comparing the different pipelines of the proposed method one can observe that, similarly to Hand and quite unlike Jumper and Tea Towel datasets, full calibration is absolutely key to accurate reconstruction of the challenging Face dataset. The pure geometries in Figure 5.24 show a dramatic improvement in both global accuracy and smoothness of reconstruction with full calibration. As previously observed on many other datasets, partial calibration actually worsens the reconstruction: observe the general degradation in smoothness moving from the uncalibrated to the partially calibrated result as well as the sharpening of local artefacts on the cheeks (e.g. frames 1 and 80). Upgrading to full calibration greatly steps up both global and local accuracy as well as smoothness compared to both the partially calibrated and uncalibrated results. Figure 5.25 exposes the superiority of the fully calibrated CtF MRF HS in the normal estimates. Specifically, incomplete (as well as lack of) calibration causes a global distortion of photometric normal orientation visualised by the erroneous (darker) shading of the reconstructions in Figure 5.25.

Although not flawless, the reconstructions obtained with the fully calibrated pipeline are very good considering that the reconstructed surface exhibits sub-surface scattering violat-
Chapter 5. CL HS for dynamic scenes with arbitrary spatially-varying chromaticity

Figure 5.24: Reconstruction of complex dynamic scenes with arbitrary spatially-varying chromaticity using CL HS: Face. Reconstruction cores compared: fully calibrated CtF ML HS, uncalibrated CtF MRF HS, partially calibrated CtF MRF HS (PhotoCalib), and only partially calibrated CtF MRF HS (proposed). Initial sampling resolution: 1 mm/0.5 mm; Final sampling resolution: 0.375 mm/0.03125 mm; Integration: NoInt shown for all with flat shading.

Frame 130  Frame 80  Frame 50  Frame 46  Frame 30  Frame 1

CtF MRF HS  CtF MRF HS  CtF MRF HS  CtF MRF HS  CtF MRF HS  CtF MRF HS
fully calibrated with PhotoCalib
uncalibrated
fully calibrated
Figure 5.25: The reconstruction results from Figure 5.24 shown in smooth shading (taking photometric normals in account).
ing the assumptions of both the chromaticity estimation procedure and Helmholtz Stereopsis in general. The algorithm even correctly reconstructs the teeth that are the uncalibrated elements of the scene, exposed in the course of the scene playing out. The teeth are characterised by a specular reflectance model and a chromaticity with an approximately equal inter-channel relationship of colour white. On the other hand, self-occlusions (e.g. by the nose) create difficult to reconstruct shadowed regions with low intensity and result in artefacts such as the bump on the right-hand side of the nose. Such self-shadowing is a common problem with acquisition set-ups consisting of a limited number of cameras.

For validation of temporal coherence a 200-frame dynamic Face sequence has been reconstructed. As before the sequence is reconstructed using the same coarse-to-fine framework but from the initial resolution of 3\text{mm}/0.5mm (spatially/depthwise) to the final resolution of 0.75\text{mm}/0.125mm instead of the finer resolutions of the individual reconstructions in Figures 5.24 and 5.25. Yet the level of detail down to the individual eye lashes even at this final resolution is substantial. Viewing the reconstruction video sequence, the frame-to-frame jitter in the results of the under-calibrated pipelines is apparent. In contrast, the sequence with full calibration is considerably more stable as can best be observed in the first portion of the sequence where the face is kept in the constant neutral pose. The global accuracy of geometry and photometric normals in the sequence follows the same trends as discussed previously for the individual frame reconstruction. Without calibration the definition of the lips is significantly worse which is understandable as those significantly deviate chromatically from the reference. The definition of the uncalibrated newly exposed areas inside the mouth are also better with full calibration although the causal link in this case is not obvious. The slight deterioration in the lip reconstruction at the end of the sequence even for the fully calibrated result is caused by the washing out of the region boundaries in the propagated chromaticity map due to error accumulation by optical flow over time, typical for tracking over longer sequences. In conclusion, the fully calibrated result is a very fair representation of the dynamic sequence featuring a photometrically complex object. It is worth mentioning that this result looks very similar in flat and smooth shading showing a high degree of consistency between depth and normal estimates very much contrary to the observation with the under-calibrated CtF MRF HS.
5.5.4 The challenges

In this section the aim is to show the performance of CL HS with CtF MRF HS on a couple of datasets strongly violating the fundamental assumptions of the technique. Sub-surface scattering/translucency and transparency are examples of optical behaviour violating the BRDF as an abstraction of reflectance at the interface of two media, the assumption of which is key in the core of Helmholtz Stereopsis (as well as any other intensity-based 3D reconstruction method). It has been shown on several datasets (Face, Jumper, Towel, Blouse etc.) that CtF MRF HS successfully copes with substantial manifestations of these phenomena but this tolerance to the violation of the BRDF assumption has got limits. Figure 5.26 shows a figurine of a turtle (“Turtle”) made of a semi-precious stone with extreme levels of translucency/sub-surface scattering as well as local saturations. The estimated chromaticity maps per camera as well as the reconstruction in flat and smooth shading with the fully calibrated CtF MRF HS are also shown.

For this object it is not straightforward to define surface chromaticity at all as the sub-surface layers with their own properties also contribute substantially to the measured reflectance. Due to sub-surface scattering the measured chromaticity will be highly directionally dependent making it impossible to estimate universally for any reconstruction configuration. Any estimate will be a function of the camera/light source positions. Saturations further corrupt the chromaticity maps (e.g. the purple dots in the viewpoint of $C_1$).

Since the object strongly defies the assumption of BRDF reflectance model, the reconstruction obtained for this object is neither as accurate nor as detailed as those of previously shown scenes in the target group. However, considering the complexity of the object the reconstruction is still acceptable: the object is recognisable and, by visual inspection, the global shape is not obviously wrong although there are severe local artefacts.

Extreme sensor saturation, even in the absence sub-surface scattering, can pose defeating challenges for the algorithm affecting the pipeline at both the chromaticity estimation stage and at reconstruction. It is important to distinguish a specular highlight from a specular saturated region. The former is a manifestation of unconstrained BRDF which Helmholtz Stereopsis is inherently equipped to deal with, while the latter indicates inadequacy in the dynamic range of the acquisition equipment translating into the loss of information on one or more channels. By regularisation Bayesian Helmholtz Stereopsis will mitigate some
Figure 5.26: Extreme sub-surface scattering challenge: Turtle. Input intensity images, estimated chromaticity (with per-pixel normalisation by the maximum chromaticity triplet component for visualisation) and the reconstructions using fully calibrated CtF MRF HS. Initial reconstruction resolution: 3mm/0.5mm (spatially/depthwise); Final reconstruction resolution: 0.1875mm/0.03125mm.
5.5. Evaluation

Figure 5.27: Extreme saturation challenge: Disney. Input intensity images for chromaticity estimation and reconstruction, per-pixel chromaticity maps, per pixel chromaticity maps aligned to the reconstructed frame (by global homography based on 20 points) and the reconstructed meshes in flat and smooth shading. Reconstruction: fully calibrated CtF MRF HS; Initial resolution: 3 mm/0.5 mm; Final resolution: 0.375 mm/0.03125 mm. Integration: NoInt.
localised saturations but it may fail if larger areas are affected particularly if the object requires per-pixel chromaticity calibration.

For illustration, consider another dataset featuring a plastic bag ("Disney") with a high frequency multi-chromatic pattern clearly requiring per-pixel calibration. In Figure 5.27 the degree of corruption by saturation of both chromaticity estimation and reconstruction frames is extreme for several cameras. The chromaticity frame corruption translates into purple patches where the signal is undefined in the chromaticity maps of $C_2$ and $C_3$. Subsequently, intensity sampling for constraint formulation at reconstruction is corrupted again by different sensor saturations. In fact, all cameras saturate in each channel somewhere in the frame, however the effect is the most pronounced for the green channel of $C_1$ manifesting itself in ubiquitous glare of that colour. With such severe signal corruption at two different stages of the pipeline it is fully expected that the algorithm will not work as confirmed by the reconstructions in Figure 5.27. This is however not a failure of the algorithm itself but rather of the acquisition equipment as for objects of such photometric complexity much higher dynamic range imagery is required.

5.6 Conclusion

Modern applications call for dynamic scene reconstruction free of limitations on the type of reflectance. The notion of a reflectance model encompasses several physical characteristics of the surface point. In traditional white light systems, it is primarily the effect of sampling geometry on the obtained response. In coloured light systems, the wavelength dependence of the reflectance model plays a role too due to the surface colour resulting in inconsistent response for different signal wavelengths. By virtue of its acquisition method Helmholtz Stereopsis is independent of the directional component of the reflectance model. However, due to sampling with multi-chromatic illumination, Colour Helmholtz Stereopsis (CL HS) introduced in the previous chapter was limited in applicability by the wavelength dependence of reflectance. In this chapter the goal was to remove this limitation and to generalise to scene reconstruction with unconstrained reflectance both directionally and in terms of response to wavelengths.

The CL HS introduced in Chapter 4 was limited to scenes with uniform chromaticity identical to the reference in the Helmholtz camera photometric calibration of the acquisition
equipment. Given such an equality to reference chromaticity, the wavelength-related response of the surface is essentially hardcoded in the photometric calibration of the equipment. With any deviation from the reference chromaticity, a conversion factor is needed to adapt the photometric calibration to the properties of the reconstructed surface by customising the wavelength dependency factor. In this chapter a statio-temporal procedure has been proposed for acquisition of such relative-to-the-reference chromaticity information whose distribution is allowed to vary spatially in each frame and temporally throughout the dynamic sequence. The introduction and seamless integration of this chromaticity calibration procedure generalised the earlier proposed CL HS to scenes with truly unconstrained reflectance.

The evaluation has revealed consistent results. One important conclusion consistently supported is that for scenes with non-reference chromaticity the procedures of photometric and chromaticity calibration are part of the same indivisible whole (“the full calibration”). For non-reference chromaticity scenes performing photometric calibration alone not only does not result in any improvement but in fact can be counter-productive (e.g. Jumper, Face, Hand). For the complex datasets with highly spatially varying and/or substantially non-reference chromaticity (e.g. Slinky, Face, Hand) the full calibration was clearly shown to be essential. Although the effects of calibration are less apparent for datasets with uniform closer to reference chromaticity (e.g. Monster, Jumper, Tea Towel), there are still measurable differences between fully calibrated and uncalibrated meshes. For closely near-reference chromaticity datasets (e.g. Tea Towel) chromaticity calibration is on the contrary not essential and the calibration may consist only of the photometric calibration of the equipment.

Further, in line with the presented theoretical discussion on reflectance behaviours deviating from the BRDF, the estimated chromaticity maps illustrate the effects of sub-surface scattering/translucency on some datasets (the fabrics, Face, Hand etc.). The effects materialise specifically in the chromaticity discrepancies between camera views. The reconstruction pipeline as a whole has been shown to cope with some level of these complex optical phenomena, despite not being intended for such surfaces: in theory both reconstruction and chromaticity estimation require direct surface reflectance exclusively. Even with extreme sub-surface scattering of Turtle, the reconstruction is not maimed beyond recognition al-
though on this dataset the artefacts are of course very apparent.

Apart from sub-surface scattering, the algorithm is sensitive to dynamic range limitations of the intensity channel responses. Saturations, such as those in the Disney dataset, can make it impossible to estimate and track chromaticity while also corrupting intensity samples for reconstruction. At the other extreme there is the pitfall of setting the camera’s black level too high in an attempt to prevent saturation which results in zero-signal on some foreground parts of the scene prohibiting their reconstruction. For reconstruction with high levels of extreme glare in the captured scene the per-channel dynamic intensity range can hence often be the bottleneck.

In summary, generalised CL HS for scenes with arbitrary spatially varying chromaticity has proven its merit showing impressively detailed reconstructions (e.g. Monster, Jumper, Towel) even for the datasets not strictly adhering in their photometric characteristics to its core principles (e.g. Face, Hand, Jumper, Towel). As for current limitations, the following can be concluded.

Firstly, the incompatibility of the method with scenes showing extreme sub-surface reflectance phenomena is fundamental as the key constraint of HS is based on the assumption of the BRDF being an accurate model of reflectance behaviour. In fact, there is currently no 3D reconstruction algorithm in existence designed to cope with scenes showing extreme levels of sub-surface scattering/translucency because of the highly non-deterministic nature of these processes. Hence, if these reflectance phenomena are prominent, they will result in reconstruction artefacts with CL HS which however have turned out to be much less pronounced than one would expect, very likely thanks to the effective tailored Bayesian HS regularisation in the core of the pipeline.

The second limitation of sensor saturation has a much larger impact on the results. The occurrence and extent of saturations can be lessened using higher dynamic range acquisition equipment, although inherently there will always be a limit to the equipment’s imaging capacity. CL HS through its Bayesian formulation is capable of mitigating moderate local saturation in input imagery acquired with limited dynamic range equipment. The signal corruption is only deadly for reconstruction if the saturations are ubiquitous and/or are also corrupting the chromaticity calibration frame for scenes with spatially-varying chromaticity (e.g. Disney).
5.6. Conclusion

The third limitation is the drift accumulated when propagating calibrated chromaticity through a long sequence resulting in the gradual washing out of the chromatic region boundaries (e.g. the Face sequence) and, as a result, local corruption of the HS constraint. This region boundary resolution loss can be addressed by replacing the classical frame-to-frame optical flow by the recent non-sequential tracking approaches or, if the sequence permits it, by repeated re-initialisation of the chromaticity map.
Chapter 5. CL HS for dynamic scenes with arbitrary spatially-varying chromaticity
Chapter 6

Conclusion and Future Work
6.1 Conclusions

The dissertation has dealt with 3D geometric reconstruction of dynamic scenes with arbitrary \textit{a priori} unknown reflectance properties. The subject matter tackles head on the challenging problem of reflectance-independent geometry acquisition, which is fundamentally different from the research concerned with reconstruction given a certain known type of reflectance, such as, on the one end of the spectrum, the simplistic Lambertian model of most established techniques or the mirror-like behaviour assumed by the strongly specular surface methods on the other end. For 3D reconstruction using intensity-based methods some knowledge of reflectance behaviour is needed to separate geometry from appearance. The advantage of Helmholtz Stereopsis over other intensity-based techniques is that it operates based on the knowledge of the generic equality of the Bidirectional Reflectance Distribution Function (BRDF) in the reciprocal sampling configuration, instead of the validity assumption of a specific parametric BRDF or some non-generic property thereof for a given surface. As long as reciprocity of sampling is guaranteed, the equality will always hold regardless of the BRDF type. In this work, the basic idea of Helmholtz Stereopsis is taken as a starting point. Through a series of contributions presented in this thesis, the technique is developed into a novel approach for the first time capable of accurate, high-resolution dynamic scene reconstruction with arbitrary \textit{a priori} unknown reflectance properties. Let us summarise the contributions.

In Chapter 3, a novel Bayesian formulation of Helmholtz Stereopsis solved as a Markov Random Field (MRF) optimisation problem is proposed, which has fundamentally improved the accuracy of geometric reconstruction. The formulation utilises neighbourhood support by means of a prior. The prior choice has been found crucial in determining the effectiveness of the formulation. A thorough study has demonstrated that priors optimising depth or normal estimates separately across the surface show well-defined biases and cannot compete with a prior enforcing consistency between depth and normal estimates. Notably such a prior is perfectly tailored to Helmholtz Stereopsis capitalising on its unique ability to generate both estimates per point. Further, the formulation of the depth-normal consistency prior is of consequence. It has been experimentally verified by a greater robustness to artefacts on many real and synthetic datasets that the
distance-based formulation is superior to the correlation-based formulation. Theoretically, the distance-based formulation has been related to integrability - a fundamental property of a vast majority of surfaces and hence a nearly universally applicable least restrictive premise for regularisation. The Bayesian formulation of Helmholtz Stereopsis is embedded into a coarse-to-fine framework resulting in high resolutions of the final point cloud. As permitted by the high resolution, the backend of the pipeline is purposely made void of any explicit integration showcasing the geometry exactly as reconstructed without running the risks of distorting the surface with post-processing. The accuracy and resolution of depth maps obtained using the proposed Coarse-to-fine Bayesian Helmholtz Stereopsis solved by MRF optimisation (CtF MRF HS) with integrability prior are shown to be far above the performance of conventional ML HS in prior art, even with the comparatively low number of eight reciprocal pairs provided as input in this chapter.

Next, in Chapter 4, Helmholtz Stereopsis is for the first time extended to dynamic scene reconstruction. To this end, Colour Helmholtz Stereopsis (CL HS) is introduced based on multi-spectral simulatenous acquisition of three reciprocal pairs. The novel multi-spectral acquisition set-up proposed for CL HS consists of three RGB cameras and three light sources of red, green and blue spectra, whose approximate pairwise collocation simulates a set-up of three multi-spectral Helmholtz cameras. The rms colocation approximation error analysed using synthetic data has been found to be confined in the range of $1 - 1.5 \text{ mm}$ and $5 - 7^\circ$ in terms of depth and normals respectively. In a multi-spectral set-up with three physically different cameras and light sources, signal consistency within reciprocal pairs becomes an issue. The problem is addressed by a practical procedure for simultaneous photometric calibration of a multi-Helmholtz-camera-pair multi-spectral configuration, generalised from its White Light Helmholtz Stereopsis (WL HS) predecessor typically performed within a single Helmholtz camera pair [106]. By incorporating CtF MRF HS without explicit integration from Chapter 3, remarkably accurate high resolution reconstructions of dynamic scenes featuring non-rigid deformation have been obtained with CL HS regardless of reflectance, even in the presence of pronounced non-Lambertian behaviour (e.g. the white laminated sheet). Although unconstrained in terms of the directional component of reflectance, the scope of CL HS in this chapter is limited in terms
of chromaticity, which is required to be uniform and matching the reference selected at the stage of photometric calibration.

In Chapter 5, CL HS is generalised to scenes with arbitrary spatially-varying chromaticity by developing a novel chromaticity calibration procedure. The procedure is essentially another stage of calibration determining the chromatic properties of the surface to be reconstructed relative to the same reference as used in the photometric calibration. By this means, the pipeline in its entirety is made reference chromaticity independent. CL HS has been evaluated on a set of scenes, both static and dynamic, with chromaticities ranging from near-reference uniform to arbitrary spatially-varying. It has been shown how chromaticity calibration becomes increasingly more important the more chromaticity deviates from the reference. Another important conclusion is the fact that the photometric and chromaticity calibration procedures are two halves of the same indivisible whole. Except for the special case of objects with uniform near-reference chromaticity, photometric calibration without chromaticity calibration will not facilitate any improvement and may actually be detrimental. The chapter also highlights the dependence of Helmholtz Stereopsis on the assumption of the BRDF being an accurate abstraction of reflectance at the surface. Following a discussion on how the BRDF violation by sub-surface reflectance phenomena affects chromaticity estimation and the basic normal constraint of Helmholtz Stereopsis, it has been shown how CL HS still produces plausible reconstructions despite corruption of data by moderate sub-surface scattering/transparency/translucency at both chromaticity estimation and reconstruction stages when sampling fabrics and human skin.

The work presented in this thesis pioneers accurate dynamic scene reconstruction without making assumptions about reflectance behaviour beyond the BRDF assumption. The acquisition set-up of CL HS does not require any highly specialised equipment beyond three RGB cameras, three projectors, three colour filters and some planar calibration targets and can be replicated in any research or industrial facility. Through its Bayesian formulation with the optimal prior related to integrability the work fundamentally improves reconstruction quality of state-of-the-art Helmholtz Stereopsis, while requiring only a minimal set of three reciprocal pairs.
6.2 Future Work

The performance of CL HS strongly depends on the accuracy of chromaticity estimation for surfaces with arbitrary spatially-varying chromaticity. Sequential propagation of chromaticity estimated at the beginning of the sequence will result in a gradual deterioration of the chromaticity map in the course of the sequence. In order to avoid this issue, non-sequential tracking techniques [126] [125] have been shown effective. Adapting non-sequential tracking to the task of chromaticity propagation to keep chromatic region boundaries sharp could facilitate reconstruction improvement further on in the sequence at the cost of computational complexity.

The dependence of CL HS on surface chromaticity can be a limiting factor of the technique in case the property cannot be effectively estimated or propagated due to, for example, extreme sensor saturations (e.g. the Disney dataset). An interesting direction to explore as future work would be the use of time-multiplexing instead of wavelength-multiplexing for simultaneous acquisition of at least the minimum number of reciprocal pairs. In the set-up the three light sources will fire sequentially within a short sampling period allowing for signal separation. In a system based on time-multiplexing of white light, surface chromaticity would no longer be an issue. However, new challenges related to scene displacement during a single sampling period will appear. Tracking techniques could once again come in useful to alleviate the problem by providing some form of motion compensation.

Both chromaticity estimation and reconstruction stages of CL HS are affected by sensor saturation resulting in locally missing intensity information. Saturation can to some extent be helped by the use of high dynamic range cameras, although there is always a limit to the effectiveness of such a hardware solution. In extreme cases saturation can completely distort the global shape of reconstruction, whereas, if the presence of saturations is localised, a saturated patch can be viewed much like surface texture. A problem arises when normal constraints get corrupted due to the intensities within the same reciprocal pair being sampled in non-corresponding textural areas. This way, texture often fools the reconstruction algorithm into generating non-existent surface structures.
These pseudo-geometries can be mitigated using any patch-based averaging techniques first proposed in [110], however these solutions are not fundamental. It would be interesting to investigate the possibility of introducing a RANSAC-based approach, as an outlier elimination stage, selecting the most consistent set of normals from the available set. Naturally, any work in this direction would require more than three reciprocal pairs to be available and hence would not be compatible with CL HS.

Finally, the proposed system based on Helmholtz Stereopsis, just as all generic intensity-based methods, inherently assumes the simplification of the more general BSSRDF of the surface to its BRDF, hence assuming direct surface reflectance only without sub-surface transport. The limiting assumption is fundamentally hardcoded in the normal constraint of the technique as Helmholtz reciprocity is characteristic of the BRDF and not of the BSSRDF. There is a fundamental difficulty in modelling sub-surface scattering phenomena because of their (partially) statistical non-deterministic nature and the dependence of every single measurement on a set of possibly heterogeneous sub-surface points not directly determined by the sampling geometry. As it cannot be modelled other than by a case-specific simplification (e.g. lumping sub-surface scattering together with diffuse reflectance in the polarisation techniques such as [127]), sub-surface scattering remains a source of error in geometry acquisition.
Bibliography


Appendices
Appendix A. Correlation-based DNprior and integrability

Appendix A contains additional derivation details of Equation 3.19 from Chapter 3 of the thesis attempting to relate the correlation-based DNprior to integrability. The normalised correlation angle $\phi_{ph-g}$ is:

$$\phi_{ph-g} = \pi^{-1} |\arcsin(n_{prj,x} \cdot t_x)|$$

where $t_x = ((\delta x)^2 + (z_2 - z_1)^2)^{-\frac{1}{2}}[\delta x, 0, (z_2 - z_1)]^\top$ and $n_{prj,x} = (n_x^2 + n_z^2)^{-\frac{1}{2}}[n_x, 0, n_z]^\top$.

The absolute value operator is introduced for consistency with the original cosine-based formulation of the correlation angle in Equation 3.7 of Chapter 3 in the $\phi_{ph-g}$ range of $[0, \frac{\pi}{2}]$.

Normals outside of the range point into the surface and can be eliminated as inconsistent prior to optimisation. Hence, one can write:

$$\left(n_{prj,1,x} \cdot t_x\right)^2 = \left(\frac{n_x \delta x + n_z (z_2 - z_1)}{n_x^2 + n_z^2}((\delta x)^2 + (z_2 - z_1)^2)\right)^\frac{1}{2} = \left((\frac{z_2 - z_1}{\delta x} + \frac{n_x}{n_z})^2 n_x^2 \delta x^2 \right)^\frac{1}{2} = \left(\frac{\delta x^2}{(\delta x)^2 + (z_2 - z_1)^2} \left(\frac{z_2 - z_1}{\delta x} - g_{x,1}\right)^2 + \frac{1}{1 + g_{x,1}^2}\right)^\frac{1}{2} = \left(\frac{z_2 - z_1}{\delta x} - g_{x,1}\right)^2 + \left(\frac{z_2 - z_1}{\delta x} g_{x,1} + 1\right)^2 = e_{Horn}^{1, dn} + \left(\frac{z_2 - z_1}{\delta x} g_{x,1} + 1\right)^2$$

The resulting relationship of the correlation-based DNprior $E_{dn}^{corr}$ to $E_{dn}^{Horn}$ is complex:

$$E_{dn}^{corr} = \frac{1}{2\pi} \arcsin \left( \frac{e_{Horn}^{1, dn}}{e_{Horn}^{1, dn} + (\frac{z_2 - z_1}{\delta x} g_{x,1} + 1)^2} \right) + \frac{1}{2\pi} \arcsin \left( \frac{e_{Horn}^{2, dn}}{e_{Horn}^{2, dn} + (\frac{z_2 - z_1}{\delta x} g_{x,2} + 1)^2} \right).$$
Appendix B. Supplementary results for Chapter 3

Appendix B presents the full set of results per each real dataset from Chapter 3. There are 8 real datasets in total in Chapter 3: teapot 1, teapot 2, doll, billiard, cup, vase, mannequin and teddy. For each dataset and each of the 5 reconstruction methods:

- standard ML HS
- Bayesian HS with Nprior
- Bayesian HS with Dprior
- Bayesian HS with corr.Dprior
- Bayesian HS with dist.DNprior (proposed)

the following results and representations are shown: the depth map, RGB normal map, flat-shaded mesh and smooth-shaded mesh (both front and side views) obtained without explicit surface integration (NoInt). In addition, front and side views of the smooth-shaded meshes obtained using the Frankot-Chellappa (FC) algorithm and Poisson surface reconstruction (PR) for integration are supplied. For objects with texture (i.e. imprinted patterns and specularities) an additional set of results is computed using patch-based averaging, instead of pointwise intensity sampling, in order to demonstrate the potential of the approach in combating pseudo-geometry reconstruction.
Real data: Teapot 1, pointwise intensity sampling (no patch-based averaging)
Real data: Teapot 1, patch-based intensity sampling
Real data: Teapot 2, pointwise intensity sampling (no patch-based averaging)

<table>
<thead>
<tr>
<th>Standard HS</th>
<th>Monte</th>
<th>Euler</th>
<th>corr. HS/HS</th>
<th>corr. Euler</th>
</tr>
</thead>
<tbody>
<tr>
<td>depth maps</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>normal maps</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>flat shading</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>smooth shading (front)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>smooth shading (side)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poisson (front)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poisson (side)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frankot-Chellappa (front)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frankot-Chellappa (side)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Real data: Teapot 2, patch-based intensity sampling
Real data: Doll, pointwise intensity sampling (no patch-based averaging)

<table>
<thead>
<tr>
<th>Method</th>
<th>Light map</th>
<th>Normal map</th>
<th>Flat shading</th>
<th>Smooth shading (front)</th>
<th>Smooth shading (side)</th>
<th>Poisson (front)</th>
<th>Poisson (side)</th>
<th>Frankot-Chellappa (front)</th>
<th>Frankot-Chellappa (side)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 depth map</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
</tr>
<tr>
<td>2 normal map</td>
<td><img src="image10.png" alt="Image" /></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
<td><img src="image15.png" alt="Image" /></td>
<td><img src="image16.png" alt="Image" /></td>
<td><img src="image17.png" alt="Image" /></td>
<td><img src="image18.png" alt="Image" /></td>
</tr>
<tr>
<td>3 flat shading</td>
<td><img src="image19.png" alt="Image" /></td>
<td><img src="image20.png" alt="Image" /></td>
<td><img src="image21.png" alt="Image" /></td>
<td><img src="image22.png" alt="Image" /></td>
<td><img src="image23.png" alt="Image" /></td>
<td><img src="image24.png" alt="Image" /></td>
<td><img src="image25.png" alt="Image" /></td>
<td><img src="image26.png" alt="Image" /></td>
<td><img src="image27.png" alt="Image" /></td>
</tr>
<tr>
<td>4 smooth shading (front)</td>
<td><img src="image28.png" alt="Image" /></td>
<td><img src="image29.png" alt="Image" /></td>
<td><img src="image30.png" alt="Image" /></td>
<td><img src="image31.png" alt="Image" /></td>
<td><img src="image32.png" alt="Image" /></td>
<td><img src="image33.png" alt="Image" /></td>
<td><img src="image34.png" alt="Image" /></td>
<td><img src="image35.png" alt="Image" /></td>
<td><img src="image36.png" alt="Image" /></td>
</tr>
<tr>
<td>5 smooth shading (side)</td>
<td><img src="image37.png" alt="Image" /></td>
<td><img src="image38.png" alt="Image" /></td>
<td><img src="image39.png" alt="Image" /></td>
<td><img src="image40.png" alt="Image" /></td>
<td><img src="image41.png" alt="Image" /></td>
<td><img src="image42.png" alt="Image" /></td>
<td><img src="image43.png" alt="Image" /></td>
<td><img src="image44.png" alt="Image" /></td>
<td><img src="image45.png" alt="Image" /></td>
</tr>
<tr>
<td>6 Poisson (front)</td>
<td><img src="image46.png" alt="Image" /></td>
<td><img src="image47.png" alt="Image" /></td>
<td><img src="image48.png" alt="Image" /></td>
<td><img src="image49.png" alt="Image" /></td>
<td><img src="image50.png" alt="Image" /></td>
<td><img src="image51.png" alt="Image" /></td>
<td><img src="image52.png" alt="Image" /></td>
<td><img src="image53.png" alt="Image" /></td>
<td><img src="image54.png" alt="Image" /></td>
</tr>
<tr>
<td>7 Poisson (side)</td>
<td><img src="image55.png" alt="Image" /></td>
<td><img src="image56.png" alt="Image" /></td>
<td><img src="image57.png" alt="Image" /></td>
<td><img src="image58.png" alt="Image" /></td>
<td><img src="image59.png" alt="Image" /></td>
<td><img src="image60.png" alt="Image" /></td>
<td><img src="image61.png" alt="Image" /></td>
<td><img src="image62.png" alt="Image" /></td>
<td><img src="image63.png" alt="Image" /></td>
</tr>
<tr>
<td>8 Frankot-Chellappa (front)</td>
<td><img src="image64.png" alt="Image" /></td>
<td><img src="image65.png" alt="Image" /></td>
<td><img src="image66.png" alt="Image" /></td>
<td><img src="image67.png" alt="Image" /></td>
<td><img src="image68.png" alt="Image" /></td>
<td><img src="image69.png" alt="Image" /></td>
<td><img src="image70.png" alt="Image" /></td>
<td><img src="image71.png" alt="Image" /></td>
<td><img src="image72.png" alt="Image" /></td>
</tr>
<tr>
<td>9 Frankot-Chellappa (side)</td>
<td><img src="image73.png" alt="Image" /></td>
<td><img src="image74.png" alt="Image" /></td>
<td><img src="image75.png" alt="Image" /></td>
<td><img src="image76.png" alt="Image" /></td>
<td><img src="image77.png" alt="Image" /></td>
<td><img src="image78.png" alt="Image" /></td>
<td><img src="image79.png" alt="Image" /></td>
<td><img src="image80.png" alt="Image" /></td>
<td><img src="image81.png" alt="Image" /></td>
</tr>
</tbody>
</table>
Real data: Billiard, pointwise intensity sampling (no patch-based averaging)

<table>
<thead>
<tr>
<th>1 depth maps</th>
<th>standard HS</th>
<th>Nprior</th>
<th>Dprior</th>
<th>corr. Dprior</th>
<th>dist. Dprior</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 normal maps</td>
<td><img src="image1" alt="Depth Maps" /></td>
<td><img src="image2" alt="Nprior" /></td>
<td><img src="image3" alt="Dprior" /></td>
<td><img src="image4" alt="corr. Dprior" /></td>
<td><img src="image5" alt="dist. Dprior" /></td>
</tr>
<tr>
<td>3 flat shading</td>
<td><img src="image6" alt="Flat Shading" /></td>
<td><img src="image7" alt="Flat Shading" /></td>
<td><img src="image8" alt="Flat Shading" /></td>
<td><img src="image9" alt="Flat Shading" /></td>
<td><img src="image10" alt="Flat Shading" /></td>
</tr>
<tr>
<td>4 smooth shading (front)</td>
<td><img src="image11" alt="Smooth Shading (Front)" /></td>
<td><img src="image12" alt="Smooth Shading (Front)" /></td>
<td><img src="image13" alt="Smooth Shading (Front)" /></td>
<td><img src="image14" alt="Smooth Shading (Front)" /></td>
<td><img src="image15" alt="Smooth Shading (Front)" /></td>
</tr>
<tr>
<td>5 smooth shading (side)</td>
<td><img src="image16" alt="Smooth Shading (Side)" /></td>
<td><img src="image17" alt="Smooth Shading (Side)" /></td>
<td><img src="image18" alt="Smooth Shading (Side)" /></td>
<td><img src="image19" alt="Smooth Shading (Side)" /></td>
<td><img src="image20" alt="Smooth Shading (Side)" /></td>
</tr>
<tr>
<td>6 Poisson (front)</td>
<td><img src="image21" alt="Poisson (Front)" /></td>
<td><img src="image22" alt="Poisson (Front)" /></td>
<td><img src="image23" alt="Poisson (Front)" /></td>
<td><img src="image24" alt="Poisson (Front)" /></td>
<td><img src="image25" alt="Poisson (Front)" /></td>
</tr>
<tr>
<td>7 Poisson (side)</td>
<td><img src="image26" alt="Poisson (Side)" /></td>
<td><img src="image27" alt="Poisson (Side)" /></td>
<td><img src="image28" alt="Poisson (Side)" /></td>
<td><img src="image29" alt="Poisson (Side)" /></td>
<td><img src="image30" alt="Poisson (Side)" /></td>
</tr>
<tr>
<td>8 Frankot-Chellappa (front)</td>
<td><img src="image31" alt="Frankot-Chellappa (Front)" /></td>
<td><img src="image32" alt="Frankot-Chellappa (Front)" /></td>
<td><img src="image33" alt="Frankot-Chellappa (Front)" /></td>
<td><img src="image34" alt="Frankot-Chellappa (Front)" /></td>
<td><img src="image35" alt="Frankot-Chellappa (Front)" /></td>
</tr>
<tr>
<td>9 Frankot-Chellappa (side)</td>
<td><img src="image36" alt="Frankot-Chellappa (Side)" /></td>
<td><img src="image37" alt="Frankot-Chellappa (Side)" /></td>
<td><img src="image38" alt="Frankot-Chellappa (Side)" /></td>
<td><img src="image39" alt="Frankot-Chellappa (Side)" /></td>
<td><img src="image40" alt="Frankot-Chellappa (Side)" /></td>
</tr>
</tbody>
</table>
Real data: Billiard, patch-based intensity sampling

<table>
<thead>
<tr>
<th>1.depth maps</th>
<th>standard BS</th>
<th>Nprior</th>
<th>Dprior</th>
<th>corr.DNprior</th>
<th>dist.DNprior</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.normal maps</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
</tr>
<tr>
<td>3.flat shading</td>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
</tr>
<tr>
<td>4.smooth shading (front)</td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
<td><img src="image15.png" alt="Image" /></td>
</tr>
<tr>
<td>5.smooth shading (side)</td>
<td><img src="image16.png" alt="Image" /></td>
<td><img src="image17.png" alt="Image" /></td>
<td><img src="image18.png" alt="Image" /></td>
<td><img src="image19.png" alt="Image" /></td>
<td><img src="image20.png" alt="Image" /></td>
</tr>
<tr>
<td>6.Poisson (front)</td>
<td><img src="image21.png" alt="Image" /></td>
<td><img src="image22.png" alt="Image" /></td>
<td><img src="image23.png" alt="Image" /></td>
<td><img src="image24.png" alt="Image" /></td>
<td><img src="image25.png" alt="Image" /></td>
</tr>
<tr>
<td>7.Poisson (side)</td>
<td><img src="image26.png" alt="Image" /></td>
<td><img src="image27.png" alt="Image" /></td>
<td><img src="image28.png" alt="Image" /></td>
<td><img src="image29.png" alt="Image" /></td>
<td><img src="image30.png" alt="Image" /></td>
</tr>
<tr>
<td>8.Frankot-Chellappa (front)</td>
<td><img src="image31.png" alt="Image" /></td>
<td><img src="image32.png" alt="Image" /></td>
<td><img src="image33.png" alt="Image" /></td>
<td><img src="image34.png" alt="Image" /></td>
<td><img src="image35.png" alt="Image" /></td>
</tr>
<tr>
<td>9.Frankot-Chellappa (side)</td>
<td><img src="image36.png" alt="Image" /></td>
<td><img src="image37.png" alt="Image" /></td>
<td><img src="image38.png" alt="Image" /></td>
<td><img src="image39.png" alt="Image" /></td>
<td><img src="image40.png" alt="Image" /></td>
</tr>
</tbody>
</table>
Real data: Cup, pointwise intensity sampling (no patch-based averaging)

<table>
<thead>
<tr>
<th>Method</th>
<th>standard HS</th>
<th>Nprior</th>
<th>Dprior</th>
<th>corr.DNprior</th>
<th>dist.DNprior</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 depth maps</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
<tr>
<td>2 normal maps</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
<tr>
<td>3 flat shading</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
<tr>
<td>4 smooth shading (front)</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
<tr>
<td>5 smooth shading (side)</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
<tr>
<td>6 Poisson (front)</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
<tr>
<td>7 Poisson (side)</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
<tr>
<td>8 Frankot-Chellappa (front)</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
</tbody>
</table>
Real data: Vase, pointwise intensity sampling (no patch-based averaging)

<table>
<thead>
<tr>
<th></th>
<th>standard HS</th>
<th>Nprior</th>
<th>Dprior</th>
<th>corr.DNprior</th>
<th>dist.DNprior</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. depth maps</td>
<td><img src="1" alt="Image" /></td>
<td><img src="2" alt="Image" /></td>
<td><img src="3" alt="Image" /></td>
<td><img src="4" alt="Image" /></td>
<td><img src="5" alt="Image" /></td>
</tr>
<tr>
<td>2. normal maps</td>
<td><img src="6" alt="Image" /></td>
<td><img src="7" alt="Image" /></td>
<td><img src="8" alt="Image" /></td>
<td><img src="9" alt="Image" /></td>
<td><img src="10" alt="Image" /></td>
</tr>
<tr>
<td>3. flat shading</td>
<td><img src="11" alt="Image" /></td>
<td><img src="12" alt="Image" /></td>
<td><img src="13" alt="Image" /></td>
<td><img src="14" alt="Image" /></td>
<td><img src="15" alt="Image" /></td>
</tr>
<tr>
<td>smooth shading (front)</td>
<td><img src="16" alt="Image" /></td>
<td><img src="17" alt="Image" /></td>
<td><img src="18" alt="Image" /></td>
<td><img src="19" alt="Image" /></td>
<td><img src="20" alt="Image" /></td>
</tr>
<tr>
<td>smooth shading (side)</td>
<td><img src="21" alt="Image" /></td>
<td><img src="22" alt="Image" /></td>
<td><img src="23" alt="Image" /></td>
<td><img src="24" alt="Image" /></td>
<td><img src="25" alt="Image" /></td>
</tr>
<tr>
<td>6. Poisson (front)</td>
<td><img src="26" alt="Image" /></td>
<td><img src="27" alt="Image" /></td>
<td><img src="28" alt="Image" /></td>
<td><img src="29" alt="Image" /></td>
<td><img src="30" alt="Image" /></td>
</tr>
<tr>
<td>7. Poisson (side)</td>
<td><img src="31" alt="Image" /></td>
<td><img src="32" alt="Image" /></td>
<td><img src="33" alt="Image" /></td>
<td><img src="34" alt="Image" /></td>
<td><img src="35" alt="Image" /></td>
</tr>
<tr>
<td>Frankot-Chellappa (front)</td>
<td><img src="36" alt="Image" /></td>
<td><img src="37" alt="Image" /></td>
<td><img src="38" alt="Image" /></td>
<td><img src="39" alt="Image" /></td>
<td><img src="40" alt="Image" /></td>
</tr>
<tr>
<td>Frankot-Chellappa (side)</td>
<td><img src="41" alt="Image" /></td>
<td><img src="42" alt="Image" /></td>
<td><img src="43" alt="Image" /></td>
<td><img src="44" alt="Image" /></td>
<td><img src="45" alt="Image" /></td>
</tr>
</tbody>
</table>
Real data: Vase, patch-based averaging

<table>
<thead>
<tr>
<th></th>
<th>standard HS</th>
<th>Nprior</th>
<th>Dprior</th>
<th>corr.DNprior</th>
<th>dist.DNprior</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 depth maps</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
</tr>
<tr>
<td>2 normal maps</td>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
</tr>
<tr>
<td>3 flat shading</td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
<td><img src="image15.png" alt="Image" /></td>
</tr>
<tr>
<td>smooth</td>
<td><img src="image16.png" alt="Image" /></td>
<td><img src="image17.png" alt="Image" /></td>
<td><img src="image18.png" alt="Image" /></td>
<td><img src="image19.png" alt="Image" /></td>
<td><img src="image20.png" alt="Image" /></td>
</tr>
<tr>
<td>smooth</td>
<td><img src="image21.png" alt="Image" /></td>
<td><img src="image22.png" alt="Image" /></td>
<td><img src="image23.png" alt="Image" /></td>
<td><img src="image24.png" alt="Image" /></td>
<td><img src="image25.png" alt="Image" /></td>
</tr>
<tr>
<td>Poisson (front)</td>
<td><img src="image26.png" alt="Image" /></td>
<td><img src="image27.png" alt="Image" /></td>
<td><img src="image28.png" alt="Image" /></td>
<td><img src="image29.png" alt="Image" /></td>
<td><img src="image30.png" alt="Image" /></td>
</tr>
<tr>
<td>Poisson (side)</td>
<td><img src="image31.png" alt="Image" /></td>
<td><img src="image32.png" alt="Image" /></td>
<td><img src="image33.png" alt="Image" /></td>
<td><img src="image34.png" alt="Image" /></td>
<td><img src="image35.png" alt="Image" /></td>
</tr>
<tr>
<td>Frankot-Chellappa (front)</td>
<td><img src="image36.png" alt="Image" /></td>
<td><img src="image37.png" alt="Image" /></td>
<td><img src="image38.png" alt="Image" /></td>
<td><img src="image39.png" alt="Image" /></td>
<td><img src="image40.png" alt="Image" /></td>
</tr>
<tr>
<td>Frankot-Chellappa (side)</td>
<td><img src="image41.png" alt="Image" /></td>
<td><img src="image42.png" alt="Image" /></td>
<td><img src="image43.png" alt="Image" /></td>
<td><img src="image44.png" alt="Image" /></td>
<td><img src="image45.png" alt="Image" /></td>
</tr>
</tbody>
</table>
Real data: Mannequin, pointwise intensity sampling (no patch-based averaging)
Real data: Teddy, pointwise intensity sampling (no patch-based averaging)