Multi view delighting and relighting

Sylvain Duchêne

UNIVERSITÉ DE NICE-SOPHIA ANTIPOLIS

ECOLE DOCTORALE STIC

SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

DISSERTATION

submitted in partial fulfillment for the degree of

DOCTOR OF SCIENCE

of the Université de Nice-Sophia Antipolis

Specialization: Computer Science

MULTI-VIEW DE-LIGHTING & RE-LIGHTING

presented by

SYLVAIN DUCHÊNE

under the supervision of George Drettakis at Inria Sophia Antipolis

April 28, 2015

Thesis committee

Reviewer	Associate Professor Dr. Diego Gutierrez	*Universidad de Zaragoza*
Reviewer	Professor Dr. Michael Goesele	*TU Darmstadt*
Examiner	Associate Professor Dr. Kartic Subr	*Heriot Watt University*
Examiner	Professor Dr. Céline Loscos	*Université Reims Champagne-Ardenne*
Examiner	Research Scientist. Dr. Adrien Bousseau	*Inria Sophia Antipolis*
Advisor	Dr. George Drettakis	*Inria Sophia Antipolis*

UNIVERSITÉ DE NICE-SOPHIA ANTIPOLIS

ECOLE DOCTORALE STIC

SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

THESE

pour l'obtention du grade de

DOCTEUR EN SCIENCES

de l'Université de Nice-Sophia Antipolis

Mention: Informatique

DÉCOMPOSITION INTRINSÈQUE MULTI-VUE & RÉ-ÉCLAIRAGE

présentée par

Sylvain DUCHÊNE

dirigée par George DRETTAKIS a Inria Sophia Antipolis

Avril 28, 2015

Jury

Rapporteur	Associate Professor Dr. Diego Gutierrez	*Universidad de Zaragoza*
Rapporteur	Professor Dr. Michael Goesele	*TU Darmstadt*
Examinateur	Associate Professor Dr. Kartic Subr	*Heriot Watt University*
Examinateur	Professor Dr. Céline Loscos	*Université Reims Champagne-Ardenne*
Examiner	Research Scientist Dr. Adrien Bousseau	*Inria Sophia Antipolis*
Directeur de thèse	Dr. George Drettakis	*Inria Sophia Antipolis*

Contents

To Arthur

Acknowledgments

My first thanks go to my advisor, George Drettakis for his natural and incredible energy. I am also grateful to Adrien Bousseau for his calm. Over the last 3 years, their deep involvement for research has always been a source of motivation for me. Being advised and working with such complementary talents was a real chance and privilege.

I owe many thanks to all my co-authors for the epic deadlines we enjoyed. I also wish to thank Reves and GraphDeco members for all these discussions that could have made us hanged or not: Fumio Okura, Abdelaziz Djelouah, Kenneth Vanhoey, Emmanuel Iarussi, Jérôme Esnault, Rodrigo Ortiz Cayon, Rahul Arora, Ayush Tewari, Charles Verron, Peter Vangorp, Joan Sol Roo, Julia Chatain, Emmanuelle Chapoulie, Laurent Lefebvre, Loïc Sévêque, Adrien David, Felicitas Hetzelt, Pierre-Yves Laffont, Jorge Lopez-Moreno, Rachid Guerchouche, Clément Riant, Christian Richardt, Gaurav Chaurasia, Stefan Popov. Of course, I do not forget to mention as well the ski sessions, beer, absinthe and whisky shared.

I also thank, Jon Barron and Vladlen Koltun for their quick and precise answers on setting up comparisons with previous work, our Autodesk collaborator Luc Robert and Emmanuel Gallo, also the Inria IT services, Francis Montagnac and Marc Vesin for providing me the required material and network resources.

I thank Diego Gutierrez, Michael Goesele, Celine Loscos and Kartic Subr for participating in my thesis committee and for spending precious time on my dissertation.

If you are a diver or you wish to become one, there is a nice diving club in Nice at the 50 Boulevard Franck Pilatte, Plongee Aigle Nautique.

Thanks to my friends for the good times, Thibault, Mathieu, Jseb, Charlotte, Amandine, Guillaume, Valérie, Mathilde, Cédric, Charline, Matthieu, Audrey, Kevin, Nicolas, Caroline et François.

My special thanks go to my oldest friends: Thomas Cordonnier, Alexis de Hautecloque, Nicolas Jouault, Ronan Lebrun and Corentin Sacré.

Finally, I wish to express my sincere thanks to Stéphanie and my family for her support along these last years.

~glhf

Abstract

We present a multi-view intrinsic decomposition algorithm that allows relighting of an outdoor scene using just a few photographs as input. Several applications such as architecture, games and movies require a 3D model of a scene. However, editing such scenes is limited by the lighting condition at the time of capture.

Our method computes intrinsic images for photos taken under the same lighting conditions with existing cast shadows by the sun. We use an automatic 3D reconstruction from these photos and the sun direction as input and decompose each image into reflectance and shading layers, despite the inaccuracies and missing data of the 3D model. Our approach is based on two key ideas.

First, we progressively improve the accuracy of the parameters of our image formation model by performing iterative estimation and combining 3D lighting simulation with 2D image optimization methods. Second, we use the image formation model to express reflectance as a function of discrete visibility values for shadow and light, which allows us to introduce a robust shadow classifier for pairs of points in a scene.

Our multi-view intrinsic decomposition is of sufficient quality for relighting of the input images. We create shadow-caster geometry which preserves shadow silhouettes and using the intrinsic layers, we can perform multi-view relighting with moving cast shadows. Our method allows image-based rendering with changing illumination conditions and reduces the cost of creating 3D content for applications.

Finally, we present an initial study on the limitation of diffuse reflectance models for these computations. We show that more complex models are required, but that simple fitting approaches are insufficient.

Résumé

Nous introduisons un algorithme de décomposition intrinsèque multi-vue qui permet de ré-éclairer une scène extérieure en utilisant quelques images en entrée. Plusieurs applications comme l'architecture, jeux et films exigent de manipuler un modèle 3D d'une scène. Cependant, la modification de telles scènes est limitée par les conditions d'éclairage de capture. Notre méthode estime les images intrinsèques pour des photos prises dans des conditions d'éclairage identiques avec des ombres. Nous utilisons conjointement une reconstruction 3D automatique et la direction du soleil pour obtenir la décomposition de chaque image en calques de réflectance et d'éclairage malgré l'inexactitude des données du modèle 3D. Notre approche est basée sur deux idées principales.

Tout d'abord, nous raffinons l'estimation des paramètres de notre modèle de formation d'image en combinant la simulation d'éclairage 3D avec des méthodes d'optimisation basée image. Deuxièmement, nous utilisons ce modèle pour exprimer la réflectance en fonction de valeur de visibilité discrète pour l'ombre et la lumière, ce qui nous permet d'introduire un classificateur d'ombre robuste pour des paires de points dans une scène. Nos calques intrinsèques sont de qualité suffisante pour manipuler les images d'entrée. Nous déplaçons les ombres portées en créant une géométrie qui préserve les silhouettes d'ombre. Notre méthode est compatible avec les approches de rendu basé image et réduit les coûts de création de contenu 3D.

Enfin, nous présentons une étude sur les limites du modèle de réflectance diffus et la difficulté d'appliquer les approches existantes dans le cadre de reconstruction 3D multi vue où les données sont imprécises.

Chapter 1

Introduction

1.1 Context and Problem Statement

Figure 1.1: Left: last sequence of Raiders of the Lost Ark (1981). Right: Jurassic Park (1993).

In 1981, Michael Pangrazio spent three months at ILM to paint this warehouse on glass to mix real footage in Raiders of the Lost Ark for the last sequence of the movie. Twelve years later, ILM created computer graphics dinosaurs and animated them with a computer less powerful than our current mobile phone as shown in Fig 1.1. Nowadays, the creation of digital content for architecture, games and movies requires photos, 3D animated models and involves experimented CG artists to mix such content. This requires a lot of time and involves the use of software dedicated to image based manipulation. Two techniques are widely used and are very popular in compositing: Rotoscoping and Matte Painting. Compositing is the high level idea of combining different sources of images into one single image. Rotoscoping refers to an image-by-image process, where the user operates a manual segmentation and matte for one part of an image or video to be used over other images. Matte Painting designates the environment creation process, from the artwork painted in the background to fine detailed digital images to create a scene.

Figure 1.2: Footage courtesy of Weta Digital available on their Youtube Channel, VFX of Dawn of the Planet of the Apes.

The development of these techniques has a direct impact on the way recent movies are produced by using green or blue screens where live footage is recorded with actors and then composited with digital content. A traditional pipeline to create 3D content is to use reference images to model objects in CG software, to paint and use a lighting transport simulation to render them. However, recent progress in automatic multi-view 3D reconstruction [Snavely *et al.*, 2006], [Goesele *et al.*, 2007], [Furukawa and Ponce, 2007] and image-based-rendering [Goesele *et al.*, 2010],[Chaurasia *et al.*, 2013] greatly facilitate the production of realistic content from a small number of photographs. Virtual walkthroughs from a small number of photographs can be achieved but still require a lot of manual interaction to get high quality sequences. Inaccurate 3D reconstruction is now part of all CG pipelines, however, multiview datasets are typically captured under fixed lighting, severely restricting their utility in games or movies – where lighting must often be manipulated to get a proper composite. Such heterogeneous content can also be developed using Nuke by The Foundry and After Effects by Adobe.

In this sequence from 2014 shown in Fig 1.2, everything is fake. Actually, only a few pictures were taken in San Francisco (a). Then, the 3D reconstructed scene was used to project texture to edit the environment, building by building, piece by piece, with the correct illumination for each of them (b). Then the artist can transform the ambience of the scene by adding 3D elements to enforce the degradation of the original scene (c), (d). Finally, the sequence is mixed with a horde of full CG apes. But this sequence highlights two main limitations: the camera motion is still very simple and the lighting conditions do not really change compared to the original scene.

We can note that the lighting conditions of the captured scene in San Francisco are similar to the final sequence which is cloudy and foggy. It is a direct consequence of one difficulty of this approach to

create content. Manipulating lighting conditions of such scene is something which cannot be achieved with existing tools. Actually removing lighting is more painful and costly than waiting for a cloudy day close enough to that required to guarantee a good composite. It avoids spending time to paint additional textures of the scene with the most plausible lighting condition thanks to the appreciation of an expert artist.

1.2 Our approach

In this thesis, we introduce an algorithm to remove lighting in such multi view datasets of outdoors scenes with cast shadows, with all photos taken in the same lighting condition, overcoming the limitation of fixed lighting in previous image-based techniques, e.g., Image-Based Rendering (IBR). We focus on wide-baseline datasets for easy capture, with a typical density of e.g., a photo per meter for a facade. Our solution decomposes each image into reflectance and shading layers, and creates a representation of movable cast shadows, allowing us to change lighting in the input images. With our approach we can plausibly modify lighting in these methods, without requiring input photos with the new illumination.

Each photograph in a multi-view dataset results from complex interactions between geometry, lighting and materials in the scene. Decomposing such images into intrinsic layers (i.e., reflectance and shading), is a hard, ill-posed problem since we have incomplete and inaccurate geometry, and lighting and materials are unknown. Previous solutions achieve impressive results for many specific subproblems, but are not necessarily adapted to automated treatment of multi-view datasets reconstructed with multi-view stereo, especially in the presence of cast shadows. For example, previous intrinsic image approaches can require manual intervention [Bousseau *et al.*, 2009], special hardware for capture [Barron and Malik, 2013a],[Chen and Koltun, 2013] or restricting assumptions on colored lighting [Garces *et al.*, 2012]. Recent learning-based shadow detectors may not always provide consistently accurate results [Guo *et al.*, 2011], and previous inverse rendering methods require pixel-accurate geometry which cannot be automatically created using multi-view stereo [Debevec *et al.*, 2004]. Our datasets can have 30-100 photographs allowing image-based navigation over a sufficient distance for image-based-rendering applications [Chaurasia *et al.*, 2013]. We thus aim for an automatic method that scales to multi-view datasets while producing consistent quality results over all views under outdoor lighting.

Our method takes the multi-view stereo 3D reconstruction as input; our algorithm is designed to handle the frequent inaccuracies and missing data of such models. The user then specifies the sun direction with two clicks, and we automatically estimate parameters of our image formation model to extract the required reflectance, shading and visibility information.

The first key idea of our approach is to progressively improve the accuracy of the image model

parameters with iterative estimation steps, by combining 3D lighting simulation with 2D image optimization.

Our second key idea is to use the image formation model to express reflectance as a function of *discrete* visibility values – 0 for shadow and 1 for light – allowing us to introduce a robust visibility classifier for pairs of points in a scene. Our method starts by finding a first estimate of sun and environment lighting parameters, as well as visibility to the sun. We then find image regions in shadow and in light, implicitly grouping regions of same reflectance. One significant difficulty of outdoor scenes is that they contain complex cast shadow boundaries. It is thus imperative to extract such boundaries as accurately as possible, which we achieve by labelling shadows using our visibility classifier. Thanks to the robustness of the shadow classifier, we can refine the estimation of the reflectance by setting constraints around shadow boundaries.

This automatic multi-view intrinsic decompositions provide high-quality layers of reflectance and shading. The quality of these decompositions is sufficient to allow us to introduce a novel approach, namely multi-view relighting with moving cast shadows. Lighting manipulation of the scene is done by manipulating the lighting layers and shadow displacements are achieved by combining the cast shadows from the inaccurate reconstructed 3D geometry and the shadow classifier. We demonstrate our approach on several multi-view datasets, and show how it can be used to achieve IBR with illumination conditions different from those of the input photos.

Our intrinsic image formation model has some limitations, but the need to more accurately model material property is one of the main limitations. To address this issue, for each 3D points we will attempt to estimate its Bidirectional Reflectance Function **BRDF** by taking as input our multi view reconstruction. Image Based BRDF measurement is not new. A pioneer of image based measurement methods is Steve Marschner. His thesis [Marschner, 1998] largely inspired the technique described in the Section 2.9. We present our results in the context of inaccurate 3D reconstruction captured in uncontrolled environment using multi view stereo techniques.

1.3 Contributions

This thesis presents the following contributions:

Lighting conditions estimation for an outdoor scene We present a method to *automatically* estimate approximate environment (indirect and sky) illumination, as well as the color of sunlight, based on lighting simulation using ray-tracing. We create an environment map representing radiance from the sky and unreconstructed objects, and combine image and 3D information to find sunlight color. Our method only takes as input a rough estimation of the sun direction set by two clicks in a minimal user interface, a multi view reconstructed 3D geometry and its input photos.

Shadow classification We introduce a shadow classifier for multi view outdoor scenes based on an intrinsic image formation model to express reflectance as a function of discrete visibility values. This approach allows us to introduce a robust visibility classifier for pairs of clusters in the scene working on the reflectance of the surface itself instead of the radiance color captured in input views. We did not address the problem of very soft shadow boundaries or fine element structures.

Intrinsic decomposition A method to compute multi-view intrinsic layers using shadow labelling and propagation. We use our robust visibility classifier in a graph labelling algorithm to assign light/shadow labels to all pixels except those in penumbra. We complete the computed intrinsic layers by propagating visibility to the remaining pixels in each image. The shadow classification is then used to improve the estimate of environment lighting, resulting in more accurate shading, visibility and reflectance layers.

Relighting Algorithm By clearly identifying regions in shadow, the quality of our intrinsic decomposition allows us to manipulate shadow regions and to re-render the scene for a new sun position by displacing the shadow and updating the lighting layer. This step is composed of four main parts: a label cleaning of the segmented regions in casted shadow, a reconstruction of a shadow caster to retrieve the original image, a shadow warping method and a shading re-evaluation. Our approach can then be combined with any image based rendering technique such as [Chaurasia *et al.*, 2013], as shown in the supplemental video.

Image Based BRDF Study The context of multi view stereo reconstruction and inverse rendering brings new challenges resulting from the inaccuracy of the 3D model and the sparsity of the captured BRDF samples. We expose these challenges and show some initial results.

1.4 Overview

The thesis is organized as follows:

- Chapter 2 presents a discussion on previous work relevant to the methods describes in this thesis.
- Chapter 3 presents a novel intrinsic decomposition algorithm based on a shadow classifier for outdoor scenes reconstructed using multi view stereo method.
- Chapter 4 introduces a novel relighting algorithm which allow the manipulation of the lighting conditions of a scene treated with the method described previously. This is the first method to provide an interactive relighting method in a multi-view setting.
- Chapter 5 describes the extensive evaluation of several steps of the algorithms introduced in Chapters 3 and 4. It also contains a ground truth estimation to study limitations of the methods.
- Chapter 6 exposes the challenge of fitting a BRDF in the context of inaccurate 3D model and presents initial results.
- Chapter 7 summarizes the results of this thesis and proposes some ideas for future work.

Chapter 2

Previous Work

The explosion of digital images over the last decade through the development and availability of DSLR cameras, mobile phones and smart phones make acquisition more and more easy. In a few seconds, anyone can capture a digital image of a sunset beach, a historic place, their garden, house and share it with friends. In this chapter we first review basics of imaging and various new approaches to image manipulation related to our work. In this thesis, we dealt with many different tools from image processing, computer vision and computer graphics. In the first three sections of this chapter, we present the basics of image capture, 3D reconstruction and global illumination, which we used as basic building blocks for our research, without actually contributing new methods. The following two sections cover methods from image processing and Markov Random Fields; our algorithms required us to adapt these solutions to our problems. The final four sections of this chapter cover previous work in intrinsic images, relighting, shadow removal and BRDF estimation which comprise the main contributions of this thesis.

2.1 Capturing an image

Digital photographs are composed of discrete color or intensity values obtained by projecting 3D geometry of a scene with specific lighting conditions, surface properties, camera optics and sensor response on a 2D plane. It is important to keep in mind that all these terms play a role in the values of an image pixel.

A camera captures the outgoing radiance L_r for a point p in one particular direction $\overrightarrow{\omega_r}$ which depends on the amount of radiance reflected from all incoming directions $\overrightarrow{\omega_i}$ through the bidirectional reflection distribution function (**BRDF**) fr as shown in Eq. (2.1).

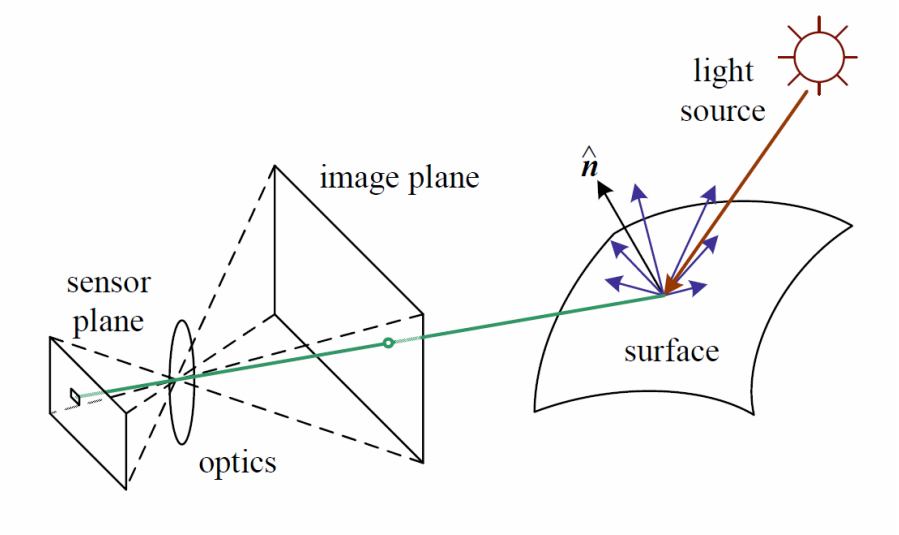


Figure 2.1: A simplified model of photometric image formation. Light is emitted by one or more light sources and is then reflected from an object's surface. A portion of this light is directed towards the camera. This model ignores multiple reflections, which occur in real-world scenes. [Szeliski, 2010a]

$$L_r(p, \vec{\omega_r}) = \int_{\Omega_i} f_r(p, \vec{\omega_i}, \vec{\omega_r}) L_i(p, \vec{\omega_i}) \cos(\theta_i) d\omega_i, \tag{2.1}$$

This observed point projects on an "imaginary" image plane of the camera (Fig.2.1) as radiance going through the lens of the camera for a particular aperture and shutter speed before finally reaching the camera CCD/CMOS sensor. The signal is then amplified through a gain stage before a standard analog to digital signal conversion (Fig. 2.2).

An image may look over or under exposed because of an incorrect choice of all camera parameters by the photographer. The shutter speed represents the exposure time and literally controls the amount of light projected to the sensor. Shutter speed is commonly also used to control motion blur effects for a waterfall, a car light in a city, objects in motion etc. The aperture corresponds to the opening camera of the diaphragm's diameter. A wide aperture will give larger depth of field by reducing the zone in focus. A smaller aperture will put the full scene in focus with no depth of field effect. Since both of those terms influence the amount of light reaching the camera sensor, one last adjustment can be performed on a capture using gain control (ISO). ISO directly refers to this legacy film sensitivity measurement still widely used for practical reasons in new digital cameras. High values give gain and too much gain also means noise. Neither the sensor nor the conversion can be controlled by the photographer, since the model of a DSLR camera defines its limits. Complex components composed of several lenses can strongly limit the range of aperture, distort images or work against chromatic aberration.

Throughout this thesis, the term image will refer to a linear image in RAW format without any post processing operation applied except the Bayer filter used to compose color images [Bayer, 1976].

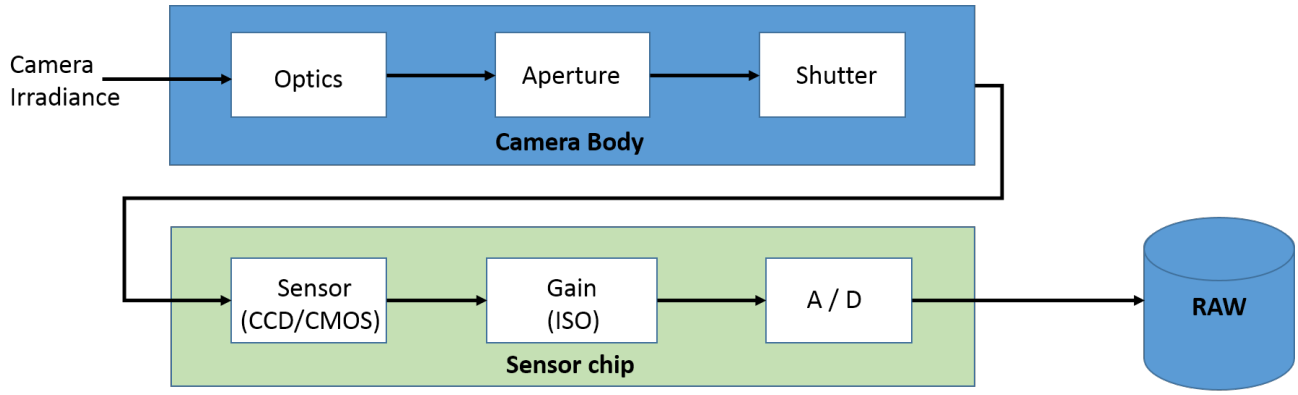


Figure 2.2: DSLR pipeline as described in [Szeliski, 2010b].

2.2 From 2D images to 3D models

Recent years have seen the new trend of producing 3D models using multiple photos of a scene for virtual visits, cinema or even 3D printing [Goesele *et al.*, 2007]. Since 3D Reconstruction is a field on its own, we will describe a traditional reconstruction pipeline. In our case we use the Photofly system from Autodesk through our partnership (Fig. 2.4). The reconstruction pipeline takes only a set of images from the same scene as input. The pipeline (Fig.2.3) contains three main steps: cameras, points and mesh reconstruction.

The first step estimates extrinsic (position in the scene) and intrinsic (focal length, radial distorsion) camera parameters. To achieve this, on each image local invariant feature detectors such as Scale-invariant feature transform (or SIFT [Lowe, 2004]) are applied to identify similar points position in multiple images. With enough samples, structure from motion algorithms can estimate 3D positions. The most popular implementation of such methods is Bundler [Snavely *et al.*, 2006]. PMVS [Furukawa and Ponce, 2007] can output a set of oriented points from the bundle adjustment, and finally a mesh can get estimated using polygonal surface or Poisson reconstruction [Kazhdan and Hoppe, 2013]. Note that this is only an example of multi view reconstruction pipeline.

To acquire a dataset for 3D reconstruction, traditional photography rules do not apply. The conditions are not the same as for a portrait in a studio with umbrellas and flashes; the camera needs to be perceived as a scanner. The goal is to maximize the quality of the 3D model, not photographic aesthetics. This notion is critical and still requires a good mastery of photography. To obtain a good reconstruction, we need pictures in focus to avoid depth of field and a good texture and lighting contrast to maximize the chance to get good feature matching.

Note all 2D to 3D methods exploit sharing information strategies between points from different input views. Many methods [Haber *et al.*, 2009], [Laffont *et al.*, 2012], [Shih *et al.*, 2013] choose to work with different lighting conditions, so they perform their acquisition by moving a light around

Figure 2.3: Overall approach from [Furukawa and Ponce, 2007]. From left to right: a sample input image; detected features; reconstructed patches after the initial matching; final patches after expansion and filtering; polygonal surface extracted from reconstructed patches.

the scene and require a tripod, or webcam video timelapse (the images overlay themselves) to avoid realignment of images or even Flickr photo collection with a risk of having images which have been manipulated and a very wide gamma-range or particular unrealistic color effects.

In our case, we want to avoid these issues and capture a scene with a single lighting condition, so it is very important to shoot using the same DSLR camera with the same lens with a fixed aperture, shutter speed and ISO. Small aperture values should be favoured to ensure the entire scene is in focus, combined with an ISO and shutter speed to maximize contrast without noise. Note that the sun can play a role when taking shots in outdoor scenes because of refraction of light rays in the lens.

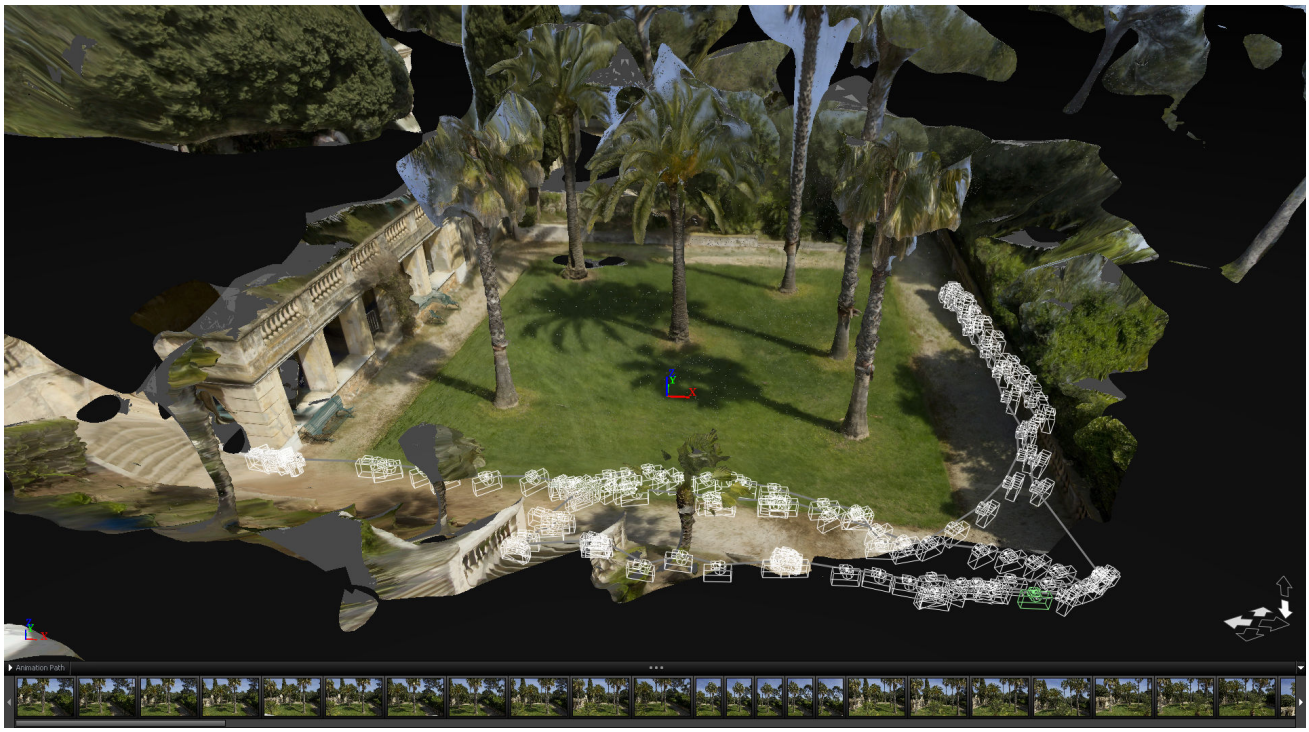


Figure 2.4: Screenshot of the photofly reconstruction of an outdoor scene showing camera locations, Refer to the dataset section to get more details.

2.3 Lighting Simulation

2.3.1 Solid Angle

The solid angle ω and its differential $d\omega$ are used in lighting simulation and computer graphics to integrate the quantity of light received or emitted by a surface Fig. 2.6. $d\Omega$ is a differential projected solid angle, representing the differential solid angle $d\omega$ with the cosine of the angle θ between the normal $\vec{n}$ of the surface and the direction $\vec{w}$ of a differential solid angle $d\omega$ Fig. 2.5).

$$d\Omega = cos\theta d\omega, \text{ with } d\omega = sin\theta d\theta d\phi \tag{2.2}$$

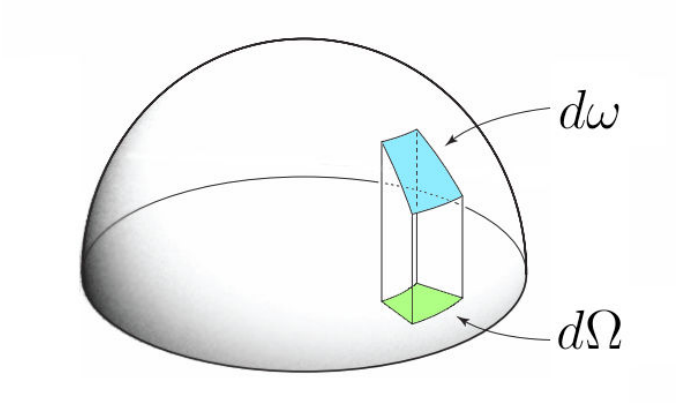


Figure 2.5: Element of solid and projected solid angle.

2.3.2 Radiometry

Radiometry is concerned with the measurement of light based on its physical properties; it should not be confused with the perceived light observed by the visual human system which refers to Photometry. Flux, irradiance/radiant exitance, intensity and radiance are the four main radiometric quantities.

Radiant flux ϕ (Eq. 2.3) measures the total energy passing through a surface during a period t in $Watt$ ($Joule.sec^{-1}$).

$$\phi = \frac{dQ}{dt} \tag{2.3}$$

Irradiance E (2.4) measures the incident flux from all incoming direction per unit surface area ($Watt.m^{-2}$). The radiant exitance B (Eq. 2.5) represents the outgoing flux per unit surface area. Both irradiance and radiant exitance have units of Watts per square meter.

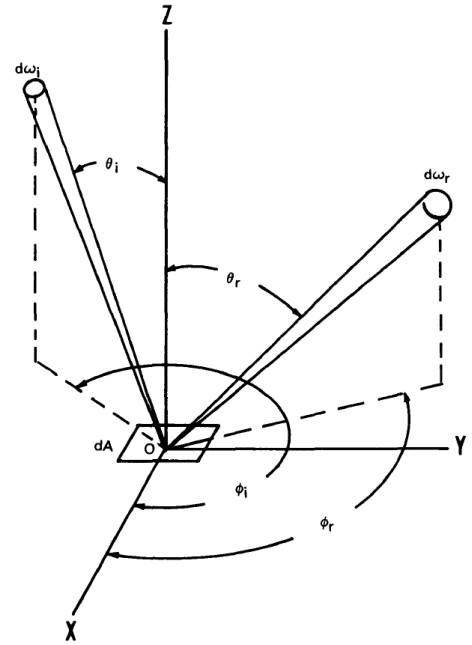


Figure 2.6: Geometry of incident and reflected elementary beams. [Nicodemus *et al.*, 1992]

$$E = \frac{d\phi_i}{dA} \tag{2.4}$$

$$B = \frac{d\phi_r}{dA} \tag{2.5}$$

The Radiant intensity I (Eq. 2.6) measures the flux per solid angle $Watt.m^{-2}$ in a direction $d\vec{w}$.

$$I = \frac{d\phi}{d\vec{w}} \tag{2.6}$$

Radiance L (Eq. 2.7)is the most important quantity. It measures the flux of light $d\phi$ in $Watt$ crossing a surface A locally perpendicular with the direction $\vec{w}$, ($\cos\theta$), as a function of its differential area dA and the solid angle dw around $\vec{w}$. It is expressed in Watts per square meter per steradian $Watt.m^{-2}.sr^{-1}$.

$$\begin{aligned} &L = \frac{d^2\phi}{dAd\omega\cos\theta} \text{with,} \\ &dA\cos\theta, \text{ the projected area,} \\ &\theta_i, \text{ the angle between the incoming light direction and the normal at the point.} \end{aligned} \tag{2.7}$$

2.3.3 The Bidirectional Reflectance Distribution

The Bidirectional Reflectance Distribution Function (**BRDF**), f_r describes the reflection of light on a surface for a given point. The incoming radiance is reflected at that same surface location x:

$$f_r(x, \overrightarrow{w_i}, \overrightarrow{w_r}) = \frac{dL_r(x, \overrightarrow{w_r})}{dE_i(x, \overrightarrow{w_i})} = \frac{dL_r((x, \overrightarrow{w_r})}{L_i(x, \overrightarrow{w_i})(\overrightarrow{n}.\overrightarrow{w_i})d\overrightarrow{w_i}}, \tag{2.8}$$

where $\overrightarrow{n}$ is the normal at point x, the vectors $\overrightarrow{w_i}$ and $\overrightarrow{w_r}$ represent the incoming and outgoing/reflected illumination direction respectively Fig. 2.6. Here, the BRDF is first defined as a ratio of reflected radiance L_r and E_i, then as the ratio between the incident and reflected differential radiance L_r which is proportional to the solid angle and $\overrightarrow{n}.\overrightarrow{w_r}$ for L_i as described in the previous section.

For a given point x, if the incident radiance field is known, the reflected radiance field can be computed; Ω is the hemisphere of incoming directions at the point x:

$$L_r(x, \overrightarrow{w_r}) = \int_\Omega f_r(x, \overrightarrow{w_i}, \overrightarrow{w_r}) dE(x, \overrightarrow{w_i}) = \int_\Omega f_r(x, \overrightarrow{w_i}, \overrightarrow{w_r}) L_i(x, \overrightarrow{w_i})(\overrightarrow{n}.\overrightarrow{w_i}) d\overrightarrow{w_i}. \tag{2.9}$$

All BRDF models have two fundamental properties which guarantee energy conservation and reciprocity. A surface modeled using a BRDF cannot produce energy and satisfies an energy conversation property which implies that the total amount of reflected light can never be bigger than the total incoming light received:

$$\int_\Omega f_r(x, \overrightarrow{w_i}, \overrightarrow{w_r})(\overrightarrow{n}.\overrightarrow{w_i}) d\overrightarrow{w_i} \leq 1, \forall \overrightarrow{w_r} \tag{2.10}$$

Helmholtz's law of reciprocity states that the **BRDF** should be symmetric and independent of the direction in which light flows, so the same result must be obtained by swapping the incoming $\overrightarrow{w_i}$ and outgoing directions $\overrightarrow{w_r}$:

$$f_r(x, \overrightarrow{w_i}, \overrightarrow{w_r}) = f_r(x, \overrightarrow{w_r}, \overrightarrow{w_i}) \tag{2.11}$$

2.3.4 Reflectance

The radiant flux incident through a solid angle $d\overrightarrow{w_i}$ onto a surface element is given by dA:

$$d\phi_i = dA \int_{\omega_i} L_i(x, \overrightarrow{w_i}) d\Omega_i \tag{2.12}$$

The flux reflected by the surface element dA into a solid angle $\overrightarrow{w_r}$ is given by:

$$d\phi_r = dA \int_{\omega_r} L_r(x, \overrightarrow{w_r}) d\Omega_r = dA \int_{\omega_r} \int_{\omega_i} f_r(x, \overrightarrow{w_i}, \overrightarrow{w_r}) L_i(x, \overrightarrow{w_i}) d\Omega_i d\Omega_r \tag{2.13}$$

The reflectance R represents the ratio of reflected to incident flux and gives the amount of light reflected by a surface at a point x:

$$R(x) = \frac{d\phi_r(x)}{d\phi_i(x)} = \frac{\int_{\omega_r} \int_{\omega_i} f_r(x, \overrightarrow{w_i}, \overrightarrow{w_r}) L_i(x, \overrightarrow{w_i}) d\Omega_i d\Omega_r}{\int_{\omega_i} L_i(x, \overrightarrow{w_i}) d\Omega_i} \tag{2.14}$$

If we assume an ideal diffuse reflection function known as Lambertian (i.e the reflected radiance is constant in all directions), we can ignore the incoming illumination direction to reflect outgoing radiance and the obtained BRDF becomes a constant $f_{r,d}$:

$$L_r(x, \overrightarrow{w_r}) = \int_{\omega_i} f_{r,d}(x, w_i, w_r) L_i d\Omega_i = f_{r,d}(x) \int_{\omega_i} L_i d\Omega_i = f_{r,d}(x) E_i(x) \tag{2.15}$$

which can be substituted directly to the reflectance $R(x)$ giving:

$$R(x) = \frac{d\phi_r(x)}{d\phi_i(x)} = \frac{dA \int_{\omega_r} \int_{\omega_i} f_r(x, \overrightarrow{w_i}, \overrightarrow{w_r}) L_i(x, \overrightarrow{w_i}) d\Omega_i d\Omega_r}{E_i(x) dA} = \frac{L_r(x) \int_{\omega_r} d\overrightarrow{w_r}}{E_i(x)} = \pi f_{r,d}(x) \tag{2.16}$$

$$\forall \overrightarrow{w_r}, L_r(x, \overrightarrow{w_r}) = L_r(x) = \frac{R(x) E_i(x)}{\pi} \tag{2.17}$$

This last relation Eq. 2.17 will be used for the intrinsic images decomposition problem described later in chapter 3, where $L_r(x, \overrightarrow{w_o})$ represents the intensity of one pixel in image, $R(x)$ the reflectance and $E_i(x)$ the shading.

2.3.5 Rendering Algorithm

The purpose of computer graphics is to produce images by simulating the reflection of light to estimate the outgoing radiance L_o as the sum of the emitted radiance L_e and the reflected radiance L_r, Eq. 2.18:

$$L_o(x, w_r) = L_e(x, w_r) + L_r(x, w_r), \text{knowing that} Lr(x, w_r) = \int_{\omega_i} f_{r,d}(x, \overrightarrow{w_i}, \overrightarrow{w_r}) L_i(\overrightarrow{n}.\overrightarrow{w_i}) d\omega_i \tag{2.18}$$

Several algorithms exist to solve this integral, but describing them is beyond the scope of this review. For all our experiments, we have used a simple Monte Carlo ray tracer [Veach, 1997].

2.4 Image processing tools

Numerical optimization techniques are used in many fields but some particular methods have became standard in image processing. This section describes image completion and image driven propagation. These methods are related to techniques we develop later in the thesis.

2.4.1 Hole Filling

To fill holes in images, the most popular tools are based on partial differential equation **PDE** and diffusion techniques such as heat flow Fig. 2.7 or based on exemplar through graph and patch correspondences Fig. 2.8. In our context we used only Perona-Malik anisotropic diffusion [Perona and Malik, 1990].

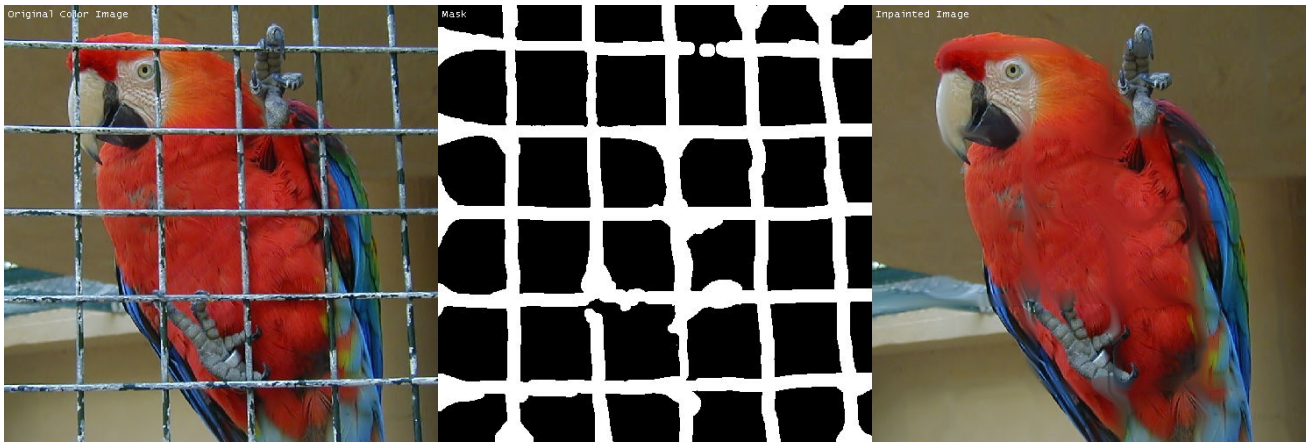


Figure 2.7: Inpainting demo of the parrot cage using anistropic diffusion. `http://cimg.sourceforge.net/greycstoration/demonstration.shtml`

Figure 2.8: Structural image editing from [Barnes *et al.*, 2009]. Left to right: (a) the original image; (b) a hole is marked (magenta) and line constraints (red/green/blue) to improve the continuity of the roofline; (c) the hole is filled in; (d) user-supplied line constraints for retargeting; (e) retargeting using constraints eliminates two columns automatically; and (f) user translates the roof upward using reshuffling.

2.4.2 Image driven Propagation

Image Driven propagation applies particularly well to the recolorization problem [Levin *et al.*, 2004] or image matting [Levin *et al.*, 2008a] to segment an image with soft boundaries instead of simple binary values. Originally developed as a matting algorithm to identify a foreground object from the background using limited user input, the method of [Levin *et al.*, 2008b] also largely inspired [Bousseau *et al.*, 2009], [Laffont *et al.*, 2012], [Laffont *et al.*, 2013] to propagate illumination (single channel or RGB channel).

Figure 2.9: Results from [Levin *et al.*, 2004], a level grey image required only some colored user strokes to propagate the color over the full image.

Figure 2.10: Image matting example using the method of [Levin *et al.*, 2008a].

2.5 Markov Random Fields

Finding a label $L \in L_0, L_1, ...L_N$ for a region R or a pixel in an image I is a popular and active research field in computer vision and image processing. Over the last years, Markov Random Fields (or MRF) have been intensively used to express inference problems given a number of measurements by using Bayes' theorem.

Bayes' Rule states that a posterior distribution $p(x|y)$ over the unknowns x given the measurements y can be obtained by multiplying the measurement likelihood $p(y|x)$ by the prior distribution $p(x)$,

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} \tag{2.19}$$

By taking the negative logarithm we get the negative posterior log likelihood. It is common to drop the constant $\log p(y)$ because its value does not matter during energy minimization since it is a normalization constant:

$$\log(p(x|y)) = -\log(p(y|x)) - \log(p(x)) + \log(p(y)) \tag{2.20}$$

The most likely or Maximum A Posteriori **MAP** solution x given y is estimated by minimizing this negative log likelihood, which can also be thought of as an energy:

$$E(x, y) = E_d(x, y) + E_p(x) \tag{2.21}$$

The term E_d is also known as the data energy or data penalty term and measures the negative log likelihood that the measurements y were observed given the unknown state x. The second term $E_p(x)$ is the prior energy also called pairwise term and it plays a role analogous to the smoothness energy in regularization. A wide range of different solutions exist to solve MRF and variants (Conditional Random Field, high-order MRF,..) under particular conditions. Gradient descent is the simplest way to optimize a MRF by updating subsets of nodes to reduce the energy configuration. However, such methods can easily get trapped in local minima requiring the use of random processes to try to get out of such minima e.g. through Markov Chain Monte Carlo updated by Gibbs sampling, stochastic gradient descent combined with simulated annealing or even linear programming relaxations (maxcut) and dynamic programming. These problems are particularly well described and illustrated in [Szeliski, 2010a]. Unfortunately, they tend to be very slow.

Such energy models are particularly useful in the context of an image modelled as a connected graph. Let us assume we have adjacent regions: R_1, R_2, R_3,...R_N to represent pixels of our image I. We want to assign a label $L \in \{s, t\}$ for each region. It is more likely that two neighbouring regions with similar colors should share the same labels. So the smoothness energy term could measure color

difference to propagate labels.

Graph Cuts [Boykov *et al.*, 2001] is a standard method to solve such energies described as a graph in Fig. 2.11 and equation (2.22) and can be applied to any pairwise MRF satisfying the sub-modular prior defined in equation (2.23).

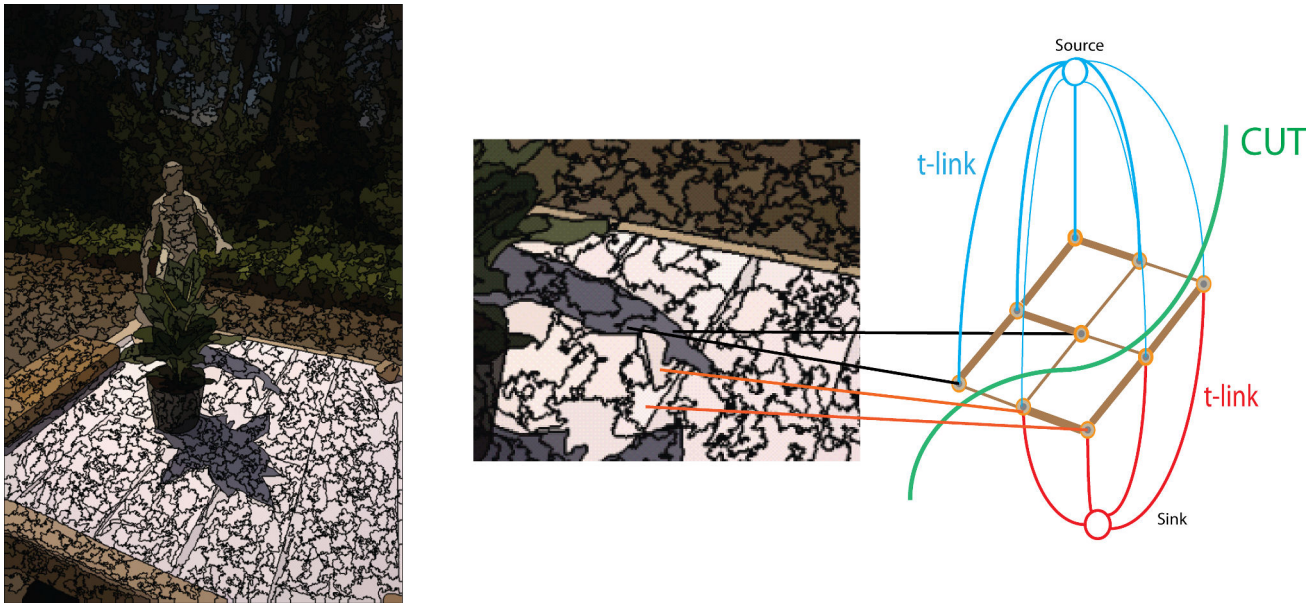


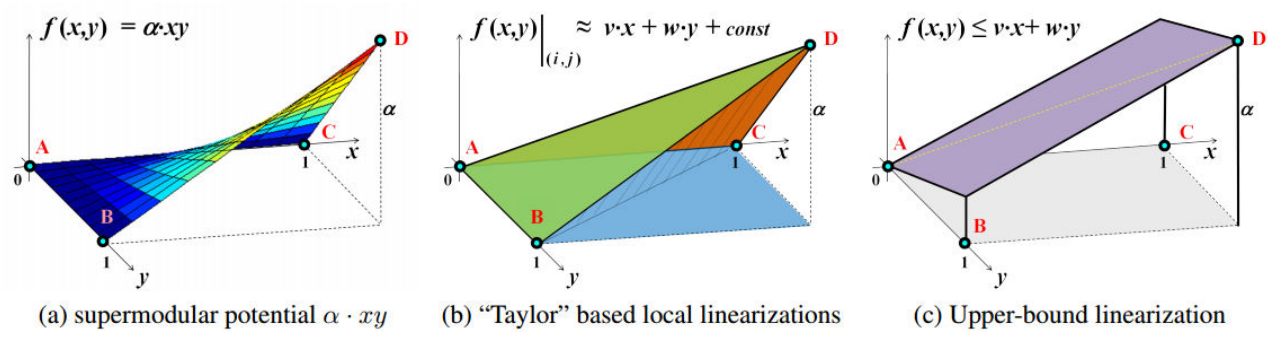
Figure 2.11: Segmentation of a clustered image. Each node in the graph corresponds to an image cluster. Edges between the nodes represents the cost of a color based affinity function which implies that similar colored cluster should have the same label, it is called the pairwise term. Each node is connected to the source and the sink to model the cost of assigning a particular label to this node, this is the data term, it can be initialized by an user interface

$$E(L) = \sum_{p} E_d(L_p) + \sum_{pq \in N} E_p(L_p, L_q), \text{with} L_p \in \{s, t\}. \tag{2.22}$$

$$E(s, s) + E(t, t) \leq E(s, t) + E(t, s) \tag{2.23}$$

However, some problems cannot be written to respect a sub-modular energy and are NP hard [Kolmogorov and Zabin, 2004]. The work on cooperative cuts by [Jegelka and Bilmes, 2011] demonstrates how to use graph cut as a subroutine. High-order MRF also introduces cost functions to express the prior between non-local and neighbouring pixels/regions which can easily break the sub-modular constraint. One way to tackle this issue is to perform inference on the graph and to refer to a message passing algorithm like Belief Propagation **BP**, also known as sum-product message [Pearl, 1982] which is mainly designed for acyclic graphs. In the case of a graph which contains cycles or loops, Loopy Belief Propagation **LBP** has been proposed. However, its convergence to obtain the maximum a posterior solution was studied later by [Weiss, 2000]. Despite the lack of a demonstration, these local

message techniques were extended by [Wainwright *et al.*, 2003] with tree-reweighted message product **TRW** and then popularized by [Kolmogorov, 2006] with the sequential tree-reweighted message product **TRW-S**. **TRW** was inspired by the problem of maximizing a lower bound on the energy and actually **TRW-S** guarantees the bound will not decrease. Binary pairwise MRF can also be optimized by the linear programming **LP** relaxation method and especially quadratic pseudo-Boolean optimization **QPBO** [Boros and Hammer, 2002] which labeled the graph partially. Recently, [Gorelick *et al.*, 2014] propose a new method to optimize non sub-modular energies outperforming **QPBO**, **LBP** and **TRW-S** methods. The key idea is to use non-linear local sub modular approximations **LSA** instead of linear approximation.



(a) supermodular potential $\alpha \cdot xy$ (b) "Taylor" based local linearizations (c) Upper-bound linearization

Figure 2.12: Figure from the [Gorelick *et al.*, 2014] method.

In Fig. 2.12, we can observe a local linearization of supermodular pairwise potential $f(x, y) = \alpha.xy$ for $\alpha > 0$. This potential defines four costs $f(0, 0) = f(0, 1) = f(1, 0) = 0$ and $f(1, 1) = \alpha$ at four distinct configurations of binary variables $x, y \in \{0, 1\}$. These costs can be plotted as four 3D points A, B, C, D in (a-c). The super-modular potential f is approximate with a linear function $v.x + w.y + const$ (plane or unary potentials). Two approximation methods are proposed, LSA-TR (trust region) which is based on Taylor expansion and LSA-AUX (auxiliary functions) on upper bounds. Both of this methods aims to control the step size during the optimization. We use TRW-S in Chapter 3 for the shadow classification.

2.6 Intrinsic Images

The problem of ”intrinsic images” was defined for the first time in 1978 by Barrow and Tenebaum [Barrow and Tenenbaum, 1978] to recover properties such as shape, reflectance and illumination from a single image. [Horn, 1989] obtained the first decomposition of an image in 1974 through Land's Retinex theory proposed in 1971 [Land and McCann, 1971] assuming that high edge discontinuities tend to be reflectance, and smooth variations are mostly due to shading. Over the last decades, a collection of methods have appeared taking advantage of new devices to capture images such as digital SLR and RGB-D cameras but also by inferring extra knowledge (scribbles, 3D information, etc) from a single image. This problem is still unsolved, and is still of interest in both computer vision and computer graphics communities.

2.6.1 Single image methods

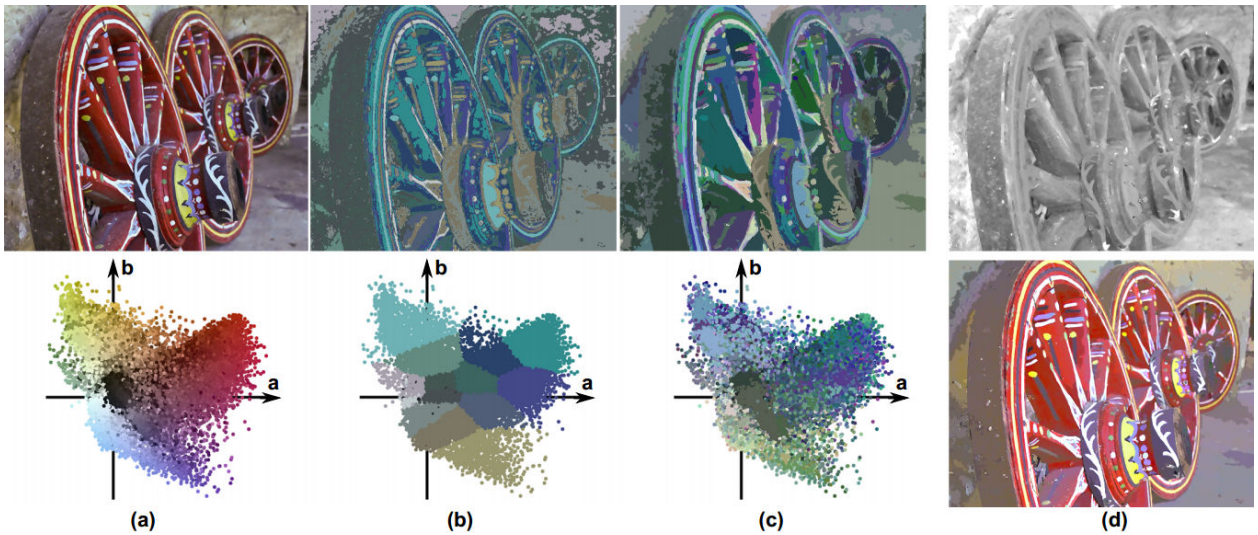


Figure 2.13: [Garces *et al.*, 2012] (a) Input image and scatter plot of pixel data in the (a,b) plane (Lab color space). (b) k-means segmentation according to (a,b) pixel coordinates. (c) Final clustering yielded by the method, taking into account spatial information (both (b) and (c) are depicted in false color). (d) The resulting shading and reflectance intrinsic images.

Many methods propose extra priors to Retinex without inferring any extra information. [Shen *et al.*, 2008] enhance Retinex priors by adding a non-local cue. They assume that for each point, there generally exists a set of other points in the image sharing the same neighbourhood texture configuration so they can enforce that such pixels share the same reflectance value. However, despite soft matching on window size and match similarity to reduce the impact of incorrect matches, some incorrect matches can remain. [Zhao *et al.*, 2012] propose a closed formula to tackle this problem solved by a simple

conjugate gradient method. [Shen *et al.*, 2011] take into account the parsimonious distribution of reflectance and assume that neighbouring pixels in a local window having similar intensity values should also have similar reflectance.

In [Garces *et al.*, 2012] Fig. 2.13, an image is considered and no additional information or user strokes are required. They propose to work with clusters instead of pixels. Clusters of similar reflectance are built based on the prior that changes in chromaticity usually correspond to changes in reflectance. They assume shading is continuous at cluster boundaries which relaxes the second Retinex assumption that shading is partially smooth. They can express this problem as a linear system representing connections and relations between clusters which can be solved quickly.

Figure 2.14: [Bousseau *et al.*, 2009] (a) Input image (b) user scribbles, white scribbles indicate fully-lit pixels, blue scribbles correspond to pixels sharing similar reflectance, red scribbles correspond to pixels sharing a similar illumination (c) (d) are respectively the obtained reflectance and illumination layer. (e) is an example of texture edition.

Since it is a largely under-constrained problem, the best results are achieved by methods using additional information. One way to provide information is to design an user interface allowing the user to partially solve the problem. [Bousseau *et al.*, 2009] propose a method to guide the decomposition using a sparse set of constraints propagated using the matting laplacian [Levin *et al.*, 2008a]. The user needs to indicate regions of constant reflectance, constant illumination or known absolute illumination (see Fig. 2.14). The main advantage of such a task is to decompose complex scenes with colored lighting or complex materials. Following the high-quality result from this work, [Carroll *et al.*, 2011] demonstrate it is also possible to recover diffuse inter-reflections. By decomposing illumination into direct lighting and indirect diffuse illumination from each material, any change in the reflectance color can be used to update indirect illumination.

With the development of machine learning to classify, model or weight, many approaches refer to such methods to take advantages of combining the best priors on local and global cues in a image and extra knowledge coming from the processing of thousands of images with ground truth. [Tappen *et al.*, 2005] combine both color information and classifier to determine if each image derivative is caused

by shading or reflectance changes. They use Belief Propagation to propagate information from areas where the correct classification is clear to areas where it is ambiguous. Following their previous work, [Tappen *et al.*, 2006] estimate intrinsic components by first estimating a set of local linear constraints, such as derivatives, from patches of the observed grayscale image. The component image is then found by solving for the image that best satisfies these constraints. They introduce a method that accounts for the uncertainty in the estimates of the constraints when solving for the estimated intrinsic component by weighting ambiguous patches. The learning process for the weight is achieved by minimizing the error between ground truth images and the their model prediction.

In their technical report, [Barron and Malik, 2013a] synthesize their work on shape illumination and reflectance from shading **SIRFS**. They pose the intrinsic decomposition problem as one statistical inference to define an optimization problem searching for the most likely explanation of a single image satisfying some reflectance, shape and illumination priors. The extra inferred knowledge is depth coming from sensor similar to a Kinect and a collection of models and weights learnt from a training set such as the one from MIT [Grosse *et al.*, 2009] and Opensurface [Bell *et al.*, 2013]. Their prior over illumination is built on a spherical-harmonic model fitted by a multivariate Gaussian to represent distant incident lighting. The prior on reflectance is composed of three assumptions:

- An assumption of piecewise consistency which is modelled by minimizing local variation.
- A parsimony of reflectance which assumes that the palette of reflectance in the scene tends to be small.
- An absolute ”reflectance” which prefers to attribute a more plausible color such as white,gray, green, brown rather than absolute black, neon pink to the scene.

The prior on shape is also composed of three assumptions:

- Smoothness, shapes tend to bend rarely.
- Isotropy of the orientation of the surface normals which reduces the fronto-parallel prior on shapes.
- Orientation of the surface normal near boundaries of masked objects.

These shape priors are imposed on intermediate representations of shape such as mean curvature or surface normals. These are computed from a depth map to satisfy all these priors and then back-propagated to the shape.

[Chen and Koltun, 2013] analyze a single RGB-D image and estimate albedo and shading fields that best explain the input. Their approach is based on the idea that the accuracy of the intrinsic decomposition can be improved if the shading image can be decomposed in different layers using lighting simulation. More particularly, they refer to direct lighting and indirect irradiance. The main advantage of referring to a physical model is the possibility to regularize each term more precisely during the minimization. They factorize an input image into four component images: an albedo, a direct irradiance

Figure 2.15: Intrinsic decomposition of an RGB-D image from the NYU Depth dataset [29]. First column Input color and depth image. Albedo and shading images estimated by two recent approaches for intrinsic decomposition of RGB-D images [Lee *et al.*, 2012], [Barron and Malik, 2013a] and by [Chen and Koltun, 2013].

image, an indirect irradiance and an illumination color image. The albedo component enforces pixels that have similar chromaticity to have similar albedo. They model illumination color as a trichromatic layer instead of a grayscale model, direct and indirect irradiance layers are modeled as scalar fields. By modeling irradiance as a 3 channel layers, the authors claim it diminished the quality of the intrinsic decomposition and only use a grayscale model. The reason is that the irradiance can change quite significantly at relatively short distances when surface curvature is high. Direct lighting regularization constrains points sharing similar normals to have similar irradiance if contributions from other objects are ignored such as shadow or inter-reflection. Indirect irradiance has to be smooth in 3D space except in regions near occlusion boundaries.

2.6.2 Multi-view and multiple lighting images methods

Recent work [Laffont *et al.*, 2013] aims to obtain intrinsic decomposition without any scan data and by limiting the acquisition tools to a DSLR camera, light probe to capture the environment map and a gray card to calibrate the sun intensity Fig. 2.16(a).The capture process does not require a laser-scan geometry or material capture as discussed in the next section [Debevec *et al.*, 2004]. Patch-based Multi-view Stereo (**PMVS**), [Furukawa and Ponce, 2010] provides a sparse reconstruction of the scene by using low dynamic range (**LDR**) input images Fig. 2.16(b). The capture of the scene also includes some high dynamic range photos (**HDR**). They infer sparse 3D lighting information using lighting simulation to compute direct lighting, sky and indirect irradiance in image space Fig. 2.16(c). Each of these terms is then propagated as a constraint using image-guided propagation [Levin *et al.*, 2004] as already demonstrated in the context of user assisted constraints [Bousseau *et al.*, 2009]. Estimating

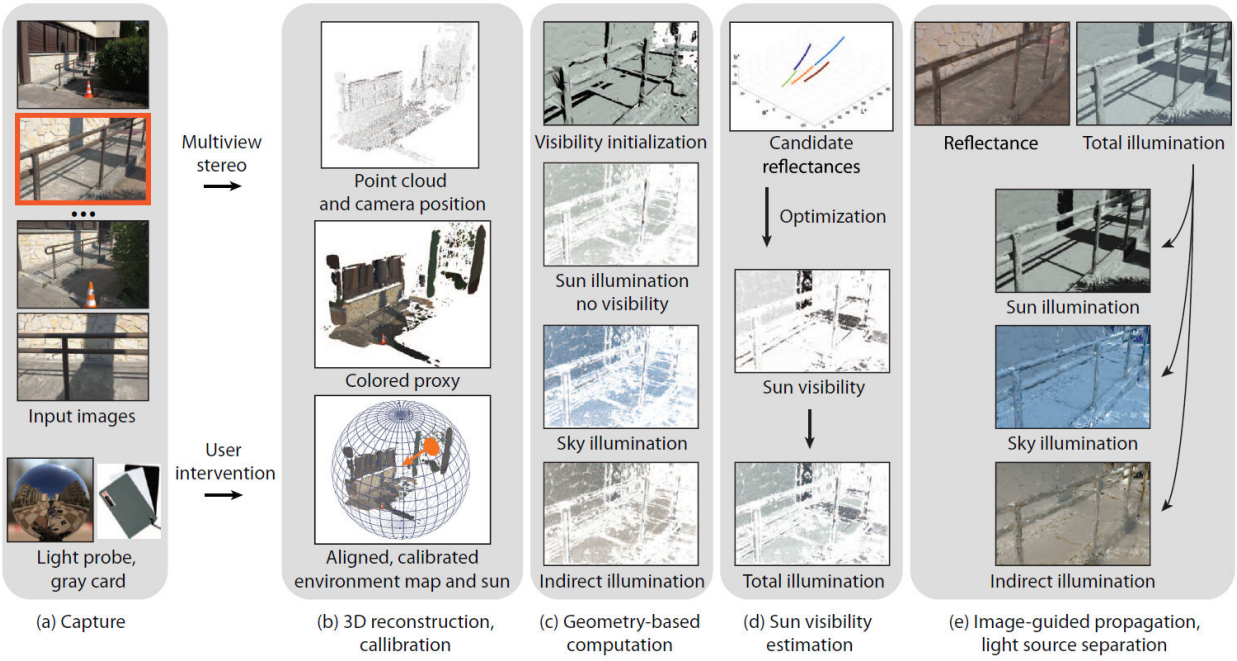


Figure 2.16: Pipeline decomposed in 5 steps [Laffont *et al.*, 2013]: (a) Capture using a DLSR camera, light probe and grey card, (b) create point cloud using [Furukawa and Ponce, 2010] then transfer radiance from images to the point cloud and aligns the scene with the environment map and sun direction, (c) lighting simulation, (d) solving sun visibility, (e) propagate lighting to obtain the reflectance.

lighting for outdoor scenes requires the classification of points in shadow, in light or in-between since the inaccurate geometry to perform a visibility test using a ray tracer is unreliable Fig. 2.16(d). The key idea is to associate pairs of clusters sharing similar reflectance with a floating visibility term. To find the best candidate, the space of solutions is explored using a mean-shift optimization to check if two clusters potentially intersect in reflectance space. However the method does not estimate mean-shift bandwidth automatically which requires user intervention to find the parameters giving the best solution.

Many approaches use the prior that reflectance remains coherent over lighting condition changes and do not necessarily require a 3D model. Through video time lapse sequences of a fixed camera, they can identify the most likely reflectance given that the scene is mostly stationary.

[Weiss, 2001] makes assumptions that derivative-like filters applied to images tend to give sparse output and claims that since filter outputs are Laplacian distributed and independent over time and space, the maximum-likelihood estimation **MLE** of the reflectance can be computed by integrating the median of the derivative filters output on a image sequence over time.

[Matsushita *et al.*, 2004a] weight the smoothness constraint on illumination by using the median estimator from derivative distributions to detect flat surfaces. Their energy model explicitly describes temporal and spatial constraints to enforce smoothness. This aims to prevent shading on non-planar

surfaces to degrade MLE in the case of adjacent pixels with different normal under a biased illumination distribution not centered around their normal (non uniform illumination conditions).[Matsushita *et al.*, 2004b] derive time-varying reflectance images and corresponding illumination images from a sequence instead of assuming a single reflectance image. Using obtained illumination images, they normalize the input image sequence in terms of incident lighting distribution to eliminate shadowing effects. This approach allows them to deal with non Lambertian scenes.

In [Sunkavalli *et al.*, 2007] a clear-sky outdoor video time lapse sequence is factorized into shadow, reflectance, illumination in sky and sun components for each pixel. This representation allows them to edit normals, shadow and reflectance. The camera viewpoint is fixed and the scene is stationary, since usually most changes in the sequence are changes in illumination. Under clear-sky assumptions lighting can be approximated as a sum of an ambient term corresponding to sky illumination and a single directional light source corresponding to the radiance of the sun. In [Sunkavalli *et al.*, 2008], they extend their method to a collection of time lapses for the same outdoor scene under several different viewpoints.

2.6.3 Evaluation

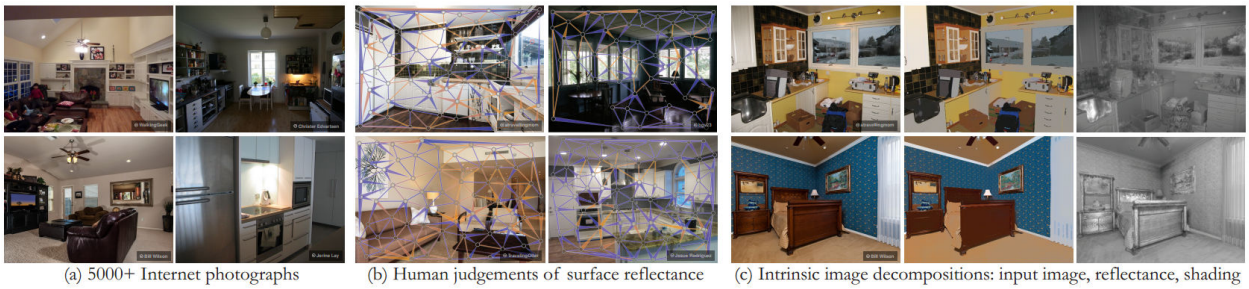


Figure 2.17: From [Bell *et al.*, 2014]. A public dataset of indoor scenes for intrinsic images in the wild is introudced (a). (b) The crowdsourcing pipeline lets users annotate pairs of points in each image with relative reflectance judgements. (c) The intrinsic image decomposition algorithm performs well in respecting the human judgements and is based on a fully-connected conditional random field (CRF) that incorporates long-range interactions in the reflectance layer while simultaneously maintaining local detail. All source images are licensed under Creative Commons

The MIT Intrinsic Images dataset [Grosse *et al.*, 2009] remains one of the best ways to evaluate the quality of an intrinsic decomposition over other approaches. To quantitatively compare several approaches, they design a dataset composed of a variety of real world objects. For each object, they provide the intrinsic decomposition of an image into 3 components: lambertian shading, reflectance and specularities. However, it focuses only a small range of materials on a small selection of single objects lit by a single direct light source. The difficulty to decompose and evaluate intrinsic decompositions for real world scenes is tackled by [Bell *et al.*, 2014] for indoors scene. They propose a new

intrinsic decomposition algorithm and an evaluation method; referring to the ability of humans to judge material comparisons despite variations in illumination. To turn it into a high scalable pipeline, they design their training in the mechanical Turk framework with a simple learning task. Instead of asking for a per pixel annotation they ask the user to disambiguate relative reflectance of pairs of pixels in each image through a simple user interface. They display an image to the user and ask them for a particular pair at the same time if cluster 1 or 2 is darker or look the same and their confidence in their evaluation. This strategy also offers a way to eliminate bad workers. Given all these judgments they introduce a metric called WHDR (Weighted Human Disagreement Rate) which measures the percentage of human judgments than an algorithm disagrees with to evaluate intrinsic decomposition. Given this, they minimize the WHDR error to optimize their energy model for each weight and parameter on a collection of variants (295 in the paper). Their energy model is designed by a dense Conditional Random Field (CRF) referring to the best priors suggested in previous work:

- Nearby pixels having similar chromaticity or intensity should have similar reflectance.
- Neighboring pixels have similar shading [Garces *et al.*, 2012]
- Reflectance is piecewise-constant [Land and McCann, 1971], [Liao *et al.*, 2013], [Barron and Malik, 2013b].
- Reflectances are sampled from a sparse set
- Certain shading values are a priori more likely than others [Barron and Malik, 2013b]
- Shading is grayscale or the same color as the light source

They demonstrate the performance of their method compared to previous methods by optimizing also each weight and parameter for all algorithms.

2.7 Inverse Rendering, Scene Manipulation, Relighting

In the literature we can find work dealing with texture editing and lighting manipulation of a scene. Intrinsic decomposition of images is mostly motivated for these reasons, however other approaches exist and each of them uses specific capture devices or requires different amount of human intervention. According to the targeted scene, indoors, outdoors, simple objects or landscape, different acquisition tools can be used ranging from a simple mobile phone camera to a 3D laser scanner.

(a) (b)

Figure 2.18: (a)Original scene.(b)Relight scene.

The work from [Loscos *et al.*, 1999] presents an interactive method to relight indoors scene (Fig. 2.18). They focus on editing real light source intensities and inserting new virtual lights and objects. Interactive updates are based on a simple model of the scene. At the time, reconstruction algorithms were not good enough to perform camera calibration without human intervention. Reconstruction was also user assisted, but the targeted scenes are easy to model with polyhedra. Hierarchical radiosity simulates lighting interaction between objects. The pipeline estimates unoccluded illumination textures which represent reflectance without taking shadows into account. Real shadows can then get reprojected via the modified radiosity estimation by modulating the texture with a ratio corresponding to the increase/decrease in illumination due to the lighting changes. Shadow boundaries are critical and require proper removal. To do so, the texture based refinement pass compares neighbouring leaves from the model decomposed in patches. If two patches have similar colors, they should share the same visibility with respect to all real light sources; also if two patches have different colors, they should have different visibility types, if not the required patch is subdivided until a decision can be taken. This way, patches around shadow boundaries will be subdivided in very fine patches and remove residuals even if ray casting failed to identify the right visibility type. All interactive updates are performed with consistent shadows of real and virtual objects.

[Debevec *et al.*, 2004] introduce a pipeline to capture a complex outdoor scene using a time of flight panoramic range 3D scanner to reconstruct the geometry mesh of the scene, chrome balls to acquire object reflectance and estimate the different environment lighting conditions such as sun direction and sun intensity. To achieve this, all acquisition devices have to be calibrated correctly which makes the preprocessing and the capture process long.

The camera is calibrated using a MacBeth color checker chart (Fig. 2.20) for each lens under natural illumination, aperture combination and multiple exposure used for data capture. Each acquired image is

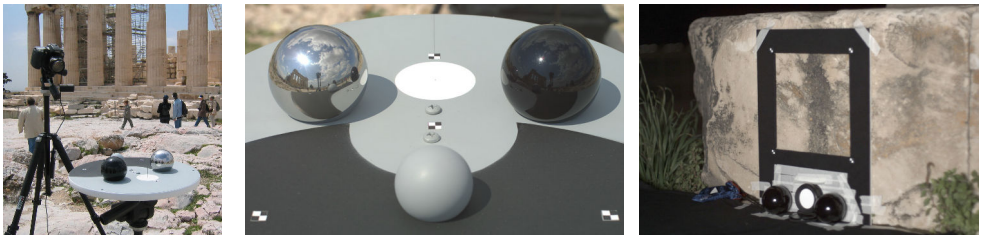


Figure 2.19: Acquisition devices use to capture BRDF.

processed to take into account this color correction before any computation. BRDF capture depends on several material representatives of an area at night-time. The region of interest is limited by a wooden square fixed by taping it on the targeted surface to capture. The setup also includes a camera, a hand-held 1000W halogen light source manipulated by a human, see Fig. 2.19 and two glossy light spheres are added over a wooden square to estimate the light source position. Some markers for the camera allows the operator to retrieve the position of the material captured in the 3D scene. One material capture process requires 40 minutes for 83 photographs.

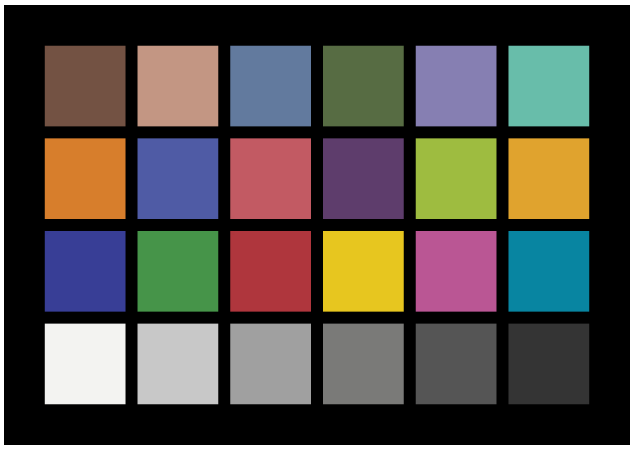


Figure 2.20: MacBeth color checker.

From this capture and a long night taking photos, some BRDF samples are obtained by picking regions of interest to manually fit a cosine lobe Lafortune BRDF model with three lobes: diffuse, specular and retro-reflective components. Based on the assumption that lambertian color is the most reliable information to match one of the BRDF samples for a surface point, the two closest BRDFs are identified by Principal Component Analysis for this component projected in a 1D space. A new BRDF is formed by interpolating specular and retroflective lobes from the Lafortune model. The idea to project in 1D space also allows materials with similar BRDFs but different color to use captured data from the material acquisition step. The environment lighting is acquired using three light probes with measured reflectance properties. Each of the spheres provides the solution for one unknown of the scene: the mirror sphere for capturing the sky and clouds, a dark shiny sphere to spot the sun and one diffuse sphere to measure its intensity see Fig. 2.19. Standard multi-bracket HDR photos are required. A semi-automatic process performs the alignment of the 3D model with pictures involving user interaction to mark correspondence of 15 points between pictures. This first estimation camera pose comes from the correspondences set of 2D pixels to 3D points. To obtain reflectance, a per image view estimation is performed and refined us-

Figure 2.21: The model rendered under novel illuminations conditions

ing multiple images until convergence for each point. Since each view will suggest different reflectance updates, the estimation is weighted by a confidence measure. The influence of each view is weighted by its angle between the view and the surface. This way photographs viewing the surface directly will have a bigger influence on the estimated reflectance properties. Also, if the estimated surface is close to an occlusion boundaries or near a strong irradiance gradient, the weight is decreased. Their algorithm starts by initializing reflectance properties obtained from BRDF estimation for all surfaces. Then for each photograph, the surfaces are rendered from this viewpoint using global illumination and the inferred Lafortune BRDF to estimate the reflectance. Multi-view reflectance information allows corrections to be applied by comparing the radiance in images at points since the Lambertian component is consistent in all views. However this pipeline introduces some limitations since it requires accurate geometry of the scene to allow rendering, materials with limited specularity since performing such updates would not necessarily converge to a good reflectance estimation and also time because most of these steps require a lot of human intervention.

Figure 2.22: [Laffont *et al.*, 2012] method's can transfer a lighting condition from one view to another.

In [Laffont *et al.*, 2012], intrinsic decomposition is estimated using images of a scene from different viewpoints and several different lighting conditions. The capture process requires only a simple DSLR camera and a multi view stereo algorithm (PMVS, VisualSFM) which produces a sparse reconstruction to obtain 3D points and normal. The core idea is to exploit multiple illumination conditions for the same point to build strong reflectance relationship between points across different location and multiple views. By assuming the scene is lambertian, the reflectance of a point does not vary between images, but the observation of this point with different lighting conditions is not enough to allow the estimation of its reflectance. However, by working with a given pair of points sharing the same normal and incoming radiance, the variation of their captured radiances from input images can only be due to reflectance. In the case where both points share the same illu-

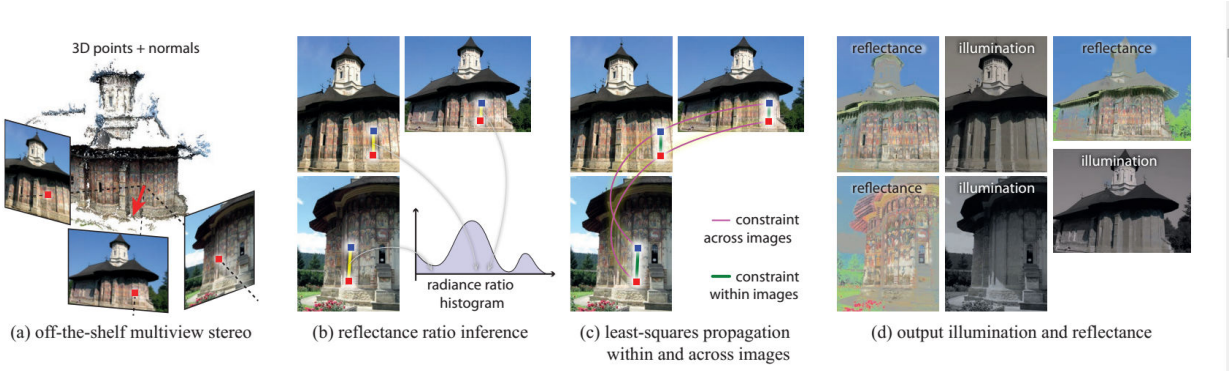


Figure 2.23: [Laffont *et al.*, 2012] method's infers reflectance ratios between points of a scene and then expresses the computation of illumination in all images in a unified least-square optimization system

mination, their ratio of radiance is equal to their ratio of reflectance. Those cases can be identified by analyzing histograms of ratios of radiance for a given pair over all images with different viewpoint and lighting. Indeed, a dominant lobe from this probability density function indicates when points of this pair receive the same incoming radiance which allows the computation of their reflectance ratio and RGB illumination. These illumination constraints are propagated through a smoothness prior already designed for intrinsic decomposition [Bousseau *et al.*, 2009], inspired by [Levin *et al.*, 2004]. When the reflectance is properly estimated, the lighting condition of one particular view can be transferred to another one see Fig 2.22. The idea to use multiple illuminations to recover an intrinsic decomposition is not new but previous methods apply it only to video time-lapse sequence as described in [Weiss, 2001], [Matsushita *et al.*, 2004a].

This aspect of lighting transfer emerges again with data driven hallucination for a single outdoor photo [Shih *et al.*, 2013]. The input data is a single outdoor photo and a collection of reference scenes captured at different times of day. The strategy to achieve lighting transfer is to match an input image with frames from a time lapse database combined with a RGB patch based mapping learnt from the time lapse. Since the most interesting lighting condition for photographers are daytime, the golden hour, blue hour and night time, each time lapse video is labelled properly to refer to those particular lighting conditions. This approach contains three steps described in (Fig 2.24). To succeed it first relies on the availability of time lapse videos similar to the input image, to reduce the set of matching videos, and this get a frame matching the time of day of the input image can be found using similar scene matching [Xiao *et al.*, 2010]. Then a dense matching is required, SIFT or SURF methods are insufficient. Methods similar to PatchMatch [Barnes *et al.*, 2009], SiftFlow [Liu *et al.*, 2008] are limited to a single image to exploit temporal coherence of the video in time lapse in their energy function. They maximize the diversity of candidate patches for a match by not selecting the best patches as the data term but by sampling randomly according to a normal law instead in the set of matched patches. This strategy improves the quality of the transfer mapping. To hallucinate the final image with the best

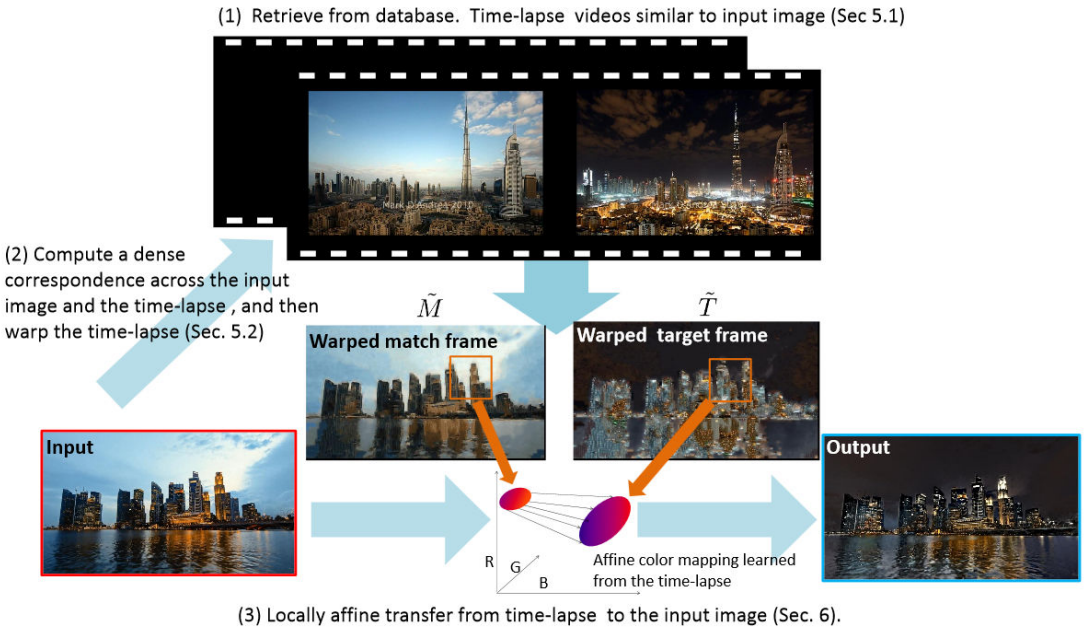


Figure 2.24: [Shih *et al.*, 2013] method's three steps. (1) They first retrieve videos of a similar scene with the input, then (2) find the local correspondence between the input and the time-lapse (image courtesy of Mark D'Andrea). (c) Finally they transfer the color appearance from the time-lapse to the input.

quality, the transfer needs to maintain the structure of the input image as well as to preserve similar color changes as seen in time lapse video. This constraint is expressed as a least squares optimization and constraints similar to Matting Laplacian [Levin *et al.*, 2008a] are introduced to guarantee data-driven propagation across the image similar to [Bousseau *et al.*, 2009]. One side effect of this mapping is noise magnification coming from the input image via capture device sensor noise or quantization. This is resolved by splitting the input image in one low frequency layer obtained by applying a bilateral filter to process the transfer and by recombining the hallucinated image with the high frequency detail image.

In both cases they achieve high-quality results because of the availability of data referring to the scene; such data are not necessarily available in the general case.

2.8 Shadow Removal and Shadow Classification

Most recent shadow detectors [Guo *et al.*, 2012], [Lalonde *et al.*, 2010], [Zhu *et al.*, 2010] are also related to intrinsic decomposition methods since their main motivation is to remove cast shadows to obtain a shadow-free image as in [Finlayson *et al.*, 2004]. Despite its importance, shadow detection remains a challenging problem. More traditional methods explore pixel or edge information. Guo *et al.* [2012] employ a region based approach to determine whether a region is shadowed compared to other

regions in the image that are likely to be of the same material. They formulate the problem as a Conditional Random Field (**CRF**), combining two terms based on a single region and a pair-wise region classification.

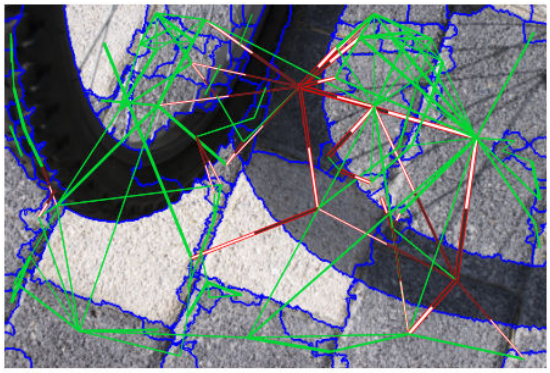


Figure 2.25: Illumination relation graph from [Guo *et al.*, 2012]. Green lines indicate same illumination pairs, red lines different illuminations pairs. White ends are the non shadow regions and gray ends are shadows.

The single region term refers to a SVM trained on manually labeled images. Their pair-wise term represents same illumination and different illumination pairs which are confidently predicted to correspond to the same material. Pairs can be built even with no adjacent regions allowing better handling of occlusion and linking similar regions divided by shadows. They identify these pairs by estimating χ^2 distances between color, texture histograms since regions sharing similar color and texture are likely to have same illumination. To guarantee oriented illumination pairs, two tests are performed:

- ratios of RGB average intensity; the non-shadow region has the highest value in both channels.
- chromatic alignment; the shadow region should not look more red or more yellow than the non-shadow region.

Their relational graph is binary with a smoothness term encouraging affinity, so their problem can be solved by Graphcut. Since shadows are not truly binary, shadow matting is estimated by using [Levin *et al.*, 2008a] and placing constraints around the shadow boundaries. Using these coefficients as a ratio between direct and environment light, they can estimated a shadow free image.

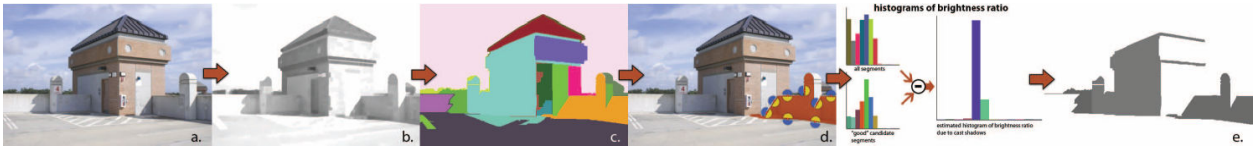


Figure 2.26: Shadow detection [Panagopoulos *et al.*, 2013] : (a) input image, (b) bright channel, (c) segmentation, (d)each segment pixel along the border is used to build a histogram of brightness ratio and identify region in shadow (e) final shadow estimate.

[Panagopoulos *et al.*, 2013] model the interaction of illumination and geometry in the scene and associate it with image evidence for cast shadows using a **MRF**. They take as input an image, approximate camera parameters and an approximate geometry model or a bounding box drawn by the user and the shape of the object is obtained using **GrabCut** [Rother *et al.*, 2004]. They have the following assumptions for the scene: lambertian reflectances, environment lighting approximated by point light sources at infinity combined with an ambient illumination term; however sky or indirect illumination

terms are not modeled. They simultaneously estimate the cast shadows, illumination and geometry parameters during the minimization and refine them at each iteration. However, estimating all these terms at the same time is very challenging, so each iteration is split in two steps. The first step estimates light and geometry parameters by selecting a candidate set parameters. Then in the second stage the focus is on the pixel labeling. The proposed **MRF** has one node for each image pixel i, one node for each light source, one node for the ambient intensity and one node for the geometry of each object in the set of objects. They penalize inconsistency between extracted shape objects and geometries, as well as for light sources which do not generate visible cast shadows in the images. They estimate shadow intensity along the boundaries of detected shadow. To initialize the optimization, they set the shadow labeled image by estimating a bright channel that guarantees and constrained to get a least 20% of image pixels fully illuminated.

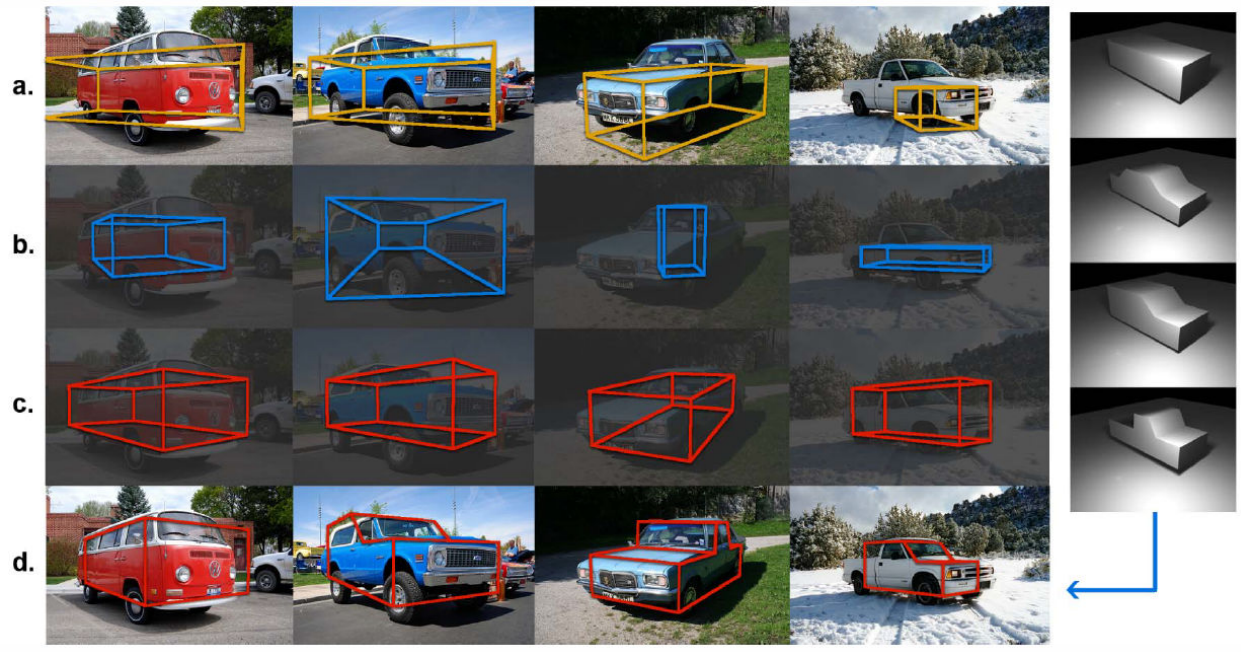


Figure 2.27: Results of joint estimation of shadows, illumination and geometry parameters [Panagopoulos *et al.*, 2013]. (a) Input image and the initial configuration of the geometry. (b) Estimated geometry by only fitting the object to the shape obtained with **GrabCut**. (c) The result with the discussed method. (d) Result using the most probable candidate geometry out of 4 classes.

2.9 Bidirectional Reflectance Distribution Function

BRDFs are 4D functions depending on the local viewing and light direction. Many BRDF models exist, the most popular are Lambertian [Lambert, 1760], Phong [Phong, 1975], Blinn-Phong [Blinn, 1977], Oren-Nayar [Oren and Nayar, 1994], Cook Torrance [Cook and Torrance, 1982]. Recovering a BRDF from an image is difficult since reflectance is confounded with shape, lighting and viewpoint. This problem is mostly tackled by inverse methods which take as input several lighting/reflectance pairs to explain any given image under a specific lighting condition. Surfaces are often curved and homogenous to generate as many samples as possible to explore the BRDF hemisphere response function. The problem is simplified in most cases by assuming isotropic BRDF. In this section we describe several methods to estimate BRDFs complementing the section on inverse rendering.

Figure 2.28: A gonioreflectometer measuring the BRDF of a planar sample by sampling the double hemisphere of incoming/outgoing radiance using a directional light source. [Li *et al.*, 2005]

A typical device to measure BRDFs is a four-axis gonioreflectometer. This device uses a combination of servo motors to position a source and a photo-detector at various locations on a hemisphere above a planar material sample. The sensor is linked to a spectroradiometer that records a measurement for each angular configuration of the source/sensor pair Fig. 2.28. The main advantage of this approach is to capture dense spectral information since a BRDF is a 5D function if we consider the wavelength dependence of the BRDF. By opposition to active reflectometry methods, passive image based approaches do not involve a gonioreflectometer but still fall in two categories: known lighting and unknown lighting conditions. Note that [Romeiro and Zickler, 2010b] synthesize their previous work [Romeiro *et al.*, 2008] and [Romeiro and Zickler, 2010a] in a technical report on inferring homogenous reflectance (BRDF) from a single HDR image of a known shape under a known and unknown real-world lighting environment. Capturing BRDFs is a very long process and can require days with

a gonioreflectometer, however acquiring them with such a device also allows the developement of a data driven approach as demonstrated by [Matusik *et al.*, 2003]. Image based BRDF measurement is mainly motivated by reducing the time of capture and also simplifying the setup required.

2.9.1 Image based methods: known lighting

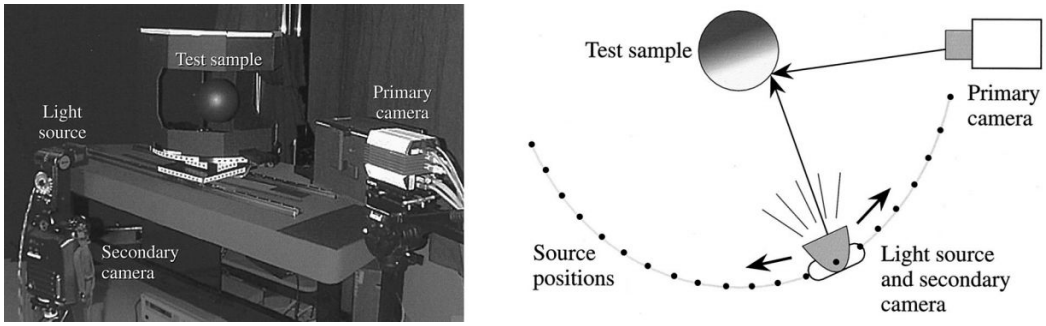


Figure 2.29: [Marschner *et al.*, 2000] setup using 2 cameras, a light source and the test sample.

[Marschner *et al.*, 2000] describe an image-based process for measuring an isotropic surface's bidirectional reflectance Fig. 2.29. Their setup includes two cameras, a uniform light source and a test sample of a known shape. They attach the second camera to the light source to measure its position using automatic photogrammetry. [Debevec *et al.*, 2004] use a similar strategy by acquiring marble sample only at night-time with a moving light probe located using 2 highly specular dark balls (Fig. 2.19).

[Matusik *et al.*, 2003] design a BRDF measurement device inspired by the image-based BRDF developed by [Marschner *et al.*, 2000] to acquire and model isotropic BRDFs.

The system takes place in a completely isolated room painted with black matte Paint and requires a spherically homogeneous sample of material. The setup includes:

- a Xenon lamp with stable emittance over the visible light range,
- a QImaging Retiga 1300 (a 10 bit, 1300x1030 resolution Firewire camera),
- a Kaidan MDT-19 (a precise computer controlled turntable)

Only the light is mounted on an arm on the turntable; the camera is stationary. The process takes about 3 hours to capture 330 HDR images with a varying exposure time ranges from 40 microseconds to 20 seconds. Their measurements process gives them typically 20-80 million BRDF samples for each material. Each pixel of the sphere is treated as a separate BRDF measurement; to do so they intersect the ray defined by the pixel with the sphere to determine the point P. Then, they compute the normal at point P on the sphere, the vector and the distance to the light source, and the vector to the camera pixel. Next, they compute the irradiance at point P due to the light source (taking into account distance to the light source and foreshortening). Finally, they can estimate the BRDF value as the ratio of the high dynamic range radiance to the irradiance. A dense set of measurements represents each BRDF, as

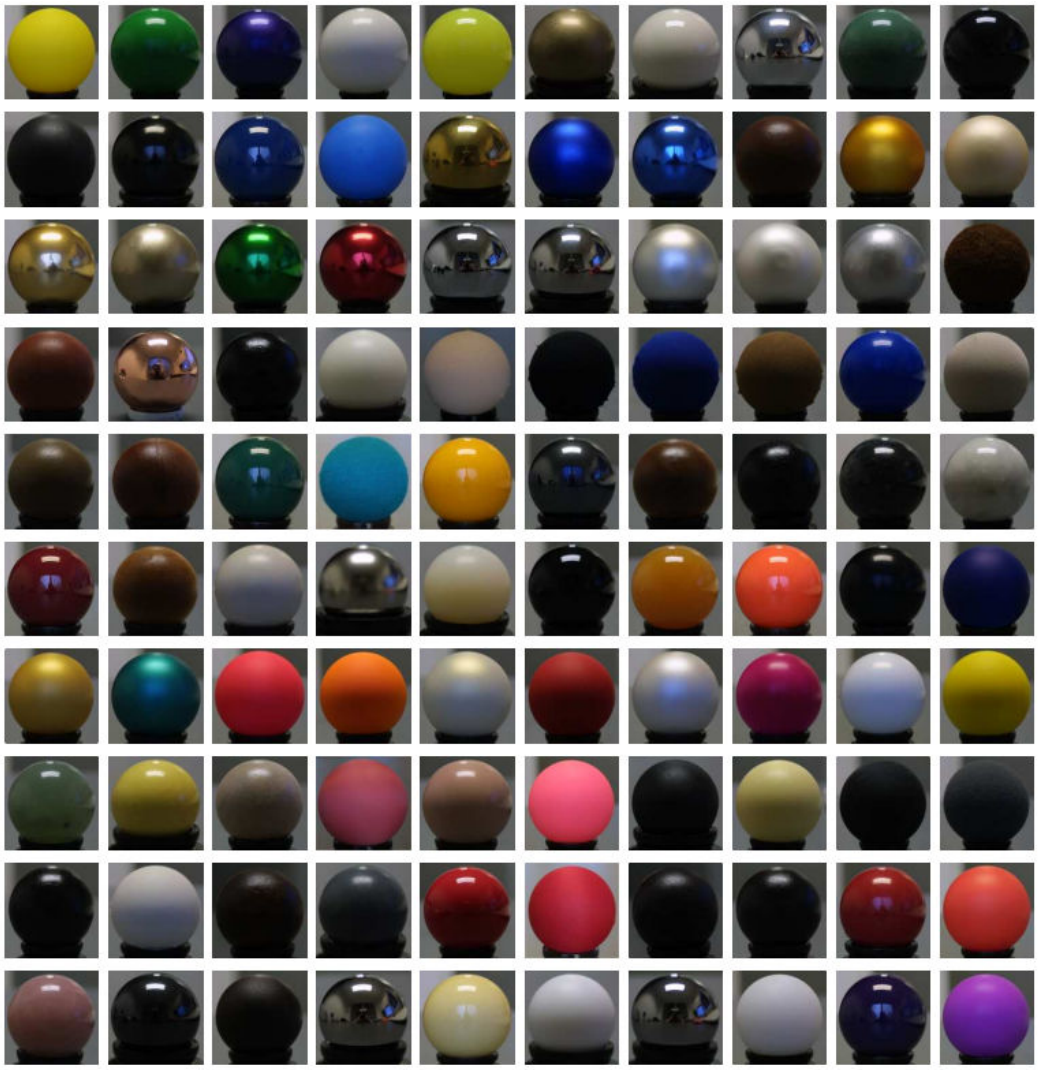


Figure 2.30: Pictures of 100 acquired materials by [Matusik *et al.*, 2003].

a table-based model which allows them to interpolate and extrapolate in the space of acquired BRDFs to create new BRDFs. However it requires the acquisition of a sufficient number of BRDFs to define a basis as shown in Fig. 2.30.

[Lensch *et al.*, 2001] and then [Lensch *et al.*, 2003] describe an acquisition process for a spatially varying BRDF which requires some manual intervention in a controlled environment Fig. 2.31. Each scene only includes a single object in a photo studio covered with dark and diffuse felt to reduce the influence of the environment composed of a single point light source, an HMI halide bulb. This bulb emits uniform light similar to a point light source. The light source position is triangulated based on captured reflections in mirroring steel balls. For each view, they acquire three sets of images: 2 images to recover the light source position, 1 image to capture the object's silhouette to register the 3D model with the images, and 1 series of photographs with bracketing to vary the exposure time

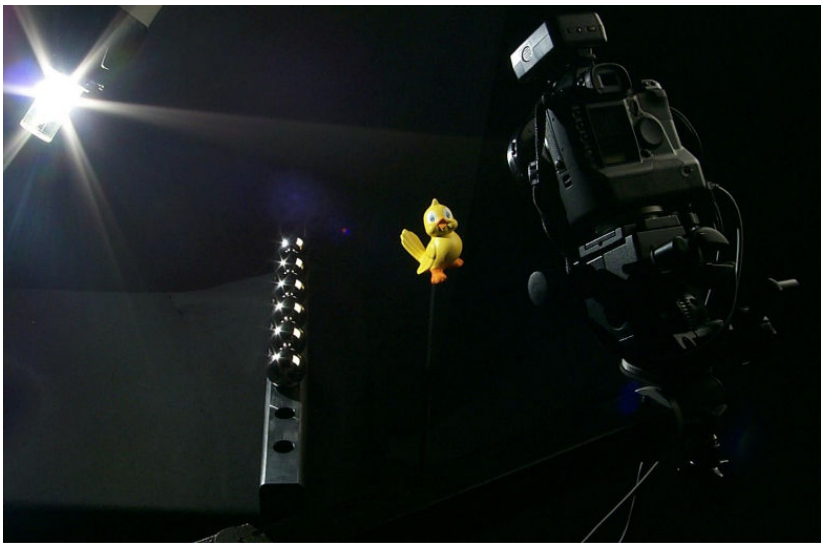


Figure 2.31: Capture setup described by [Lensch *et al.*, 2003].

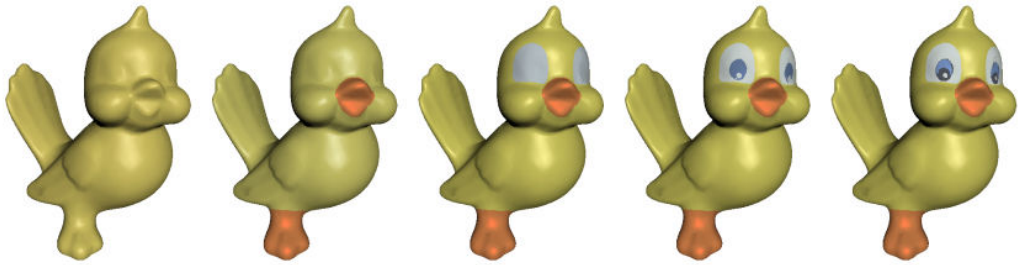


Figure 2.32: In each image, a new cluster was created using the method of [Lensch *et al.*, 2003]. The object is shaded using only the single BRDFs fit to each cluster.

and produce a HDR image. They require 20-25 HDR images for one object and obtain a 3D model using a structured light 3D scanner or a computer tomography scanner. The reconstructed mesh is then manually cleaned and decimated. They introduce the **Lumitexel** which is a very sparse representation of the BRDF for every visible surface point. This notation represents the geometric (position, normal) and photometric data of one point, the list of radiance samples R_i representing a triplet of the outgoing radiance R acquired from the surface point of the input HDR image, the direction of light and the viewing direction. Their BRDF recovery process detects the different materials of the captured object and a BRDF for all of them. The key idea is that a Lumitexel cannot be fitted with a BRDF but a group of Lumitexels in a cluster belonging to the same material can lead to a good estimation. To do this they define an error between a given BRDF and a Lumitexel which is also used for fitting and splitting BRDF cluster, see Fig. 2.32. Note that the splitting process stops based on the user estimation of the number of input materials.

The BRDF fitting involves a nonlinear optimizing method known as Levenberg-Marquadt [Marquardt, 1963] using a Lafortune BRDF model which can support up to 3 kind of lobes: diffuse, retrore-

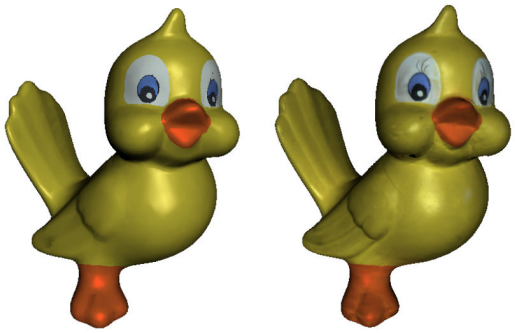


Figure 2.33: Left: The result of the clustering process does not look realistic since there is no variation of the material within one cluster. Right: Spatial variation derived by projection of the reflectance samples of each lumitexel in a basis formed by the clustered material [Lensch *et al.*, 2003].

flective and specular. To obtain a truly spatially varying BRDFs, local changes are then modeled by projecting the measured data for each surface point into a set of basis of the BRDF. The optimal set is obtained by performing a principal function analysis **PFA** using again the non linear Levenberg-Marquadt solver over a basis BRDF composed of: the BRDF which fits to the cluster f_C, the BRDFs of spatially neighboring clusters to match lumitexels at cluster boundaries f_N, the BRDF of similar clusters to the material f_M, and 2 BRDFs based on f_C, one with slightly increased and one with slightly decreased diffuse component p_d and exponent N. Finally, for each lumitexel the weights are optimized between all of them using a least square system. Note in Fig. 2.33 the difference before and after the re-projection in the basis set of BRDFs. This pipeline demonstrates its efficiency but it still requires many samples. It is important to keep in mind these figures for the bird object show in Fig. (2.31, 2.32, 2.33), **25 input views** are used, **1 917 043 lumitexels** are acquired with an average number of reflectance samples of 6.3 per lumitexel considering 5 materials clusters and **4 BRDFs bases per cluster**.

[Romeiro *et al.*, 2008] present an image based system for inferring bi-directional surface reflectance without active lighting. Their method assumes isotropic reflectance and ignores global illumination effects. It still requires a light probe, a camera, and one HDR image of a known curved homogeneous surface. Their goal is to recover general reflectance without state of the art restrictions of radial-symmetry and low-parameter models but they still assume isotropic reflectance and ignore global illumination effects. In [Alldrin and Kriegman, 2007], it is noted that some authors consider isotropy and bilateral symmetry to be distinct phenomena. A radially-symmetric BRDF is one that, like the Phong model, is radially symmetric about the reflection vector [Romeiro *et al.*, 2008]. Isotropic BRDFs are symmetric about the plane spanned by the viewing direction and surface normal. Bilaterally symmetric BRDFs can be described by the fact that the exitant radiance emitted from a bilaterally symmetric surface patch is constant when the surface is reflected about any plane collinear with its normal. [Alldrin and Kriegman, 2007] do not make such a distinction since all or nearly all physically valid isotropic BRDFs have this property. Isotropy usually refers to both symmetry and bilateral symmetry; this assumption is also

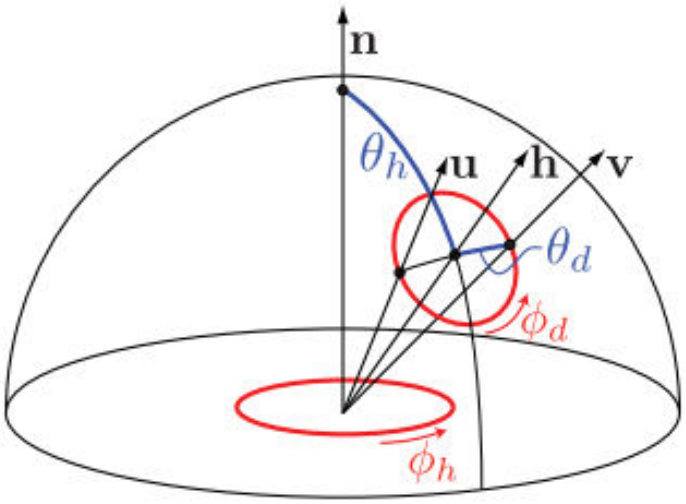


Figure 2.34: Domain reduction for reciprocal, isotropic, bilaterally-symmetric, and bivariate BRDFs. [Romeiro *et al.*, 2008] consider bivariate BRDFs, which are constant functions of ϕ_d. Isotropic BRDFs are unchanged by rotations about the surface normal (i.e., changes in ϕ_h), while reciprocity and bilateral symmetry impose periodicity for rotations about the halfway vector (i.e., changes in ϕ_d).

considered by [Romeiro *et al.*, 2008].

[Romeiro *et al.*, 2008] consider bi-variate BRDFs, which are constant functions of ϕ_d, the halfway vector and unchanged by rotations about the surface normal ϕ_f see Fig. 2.34. Their motivation to represent BRDFs with a bi-variate representation is based on the work of [Stark *et al.*, 2005] who demonstrate that a carefully selected 2D domain is often sufficient for capturing (off)-specular reflections, retro-reflections and important Fresnel effects. In the work of [Romeiro *et al.*, 2008], they design a sampling scheme which increases the number of samples near specular reflections, to use the same sampling strategy independently of the sampled material and to be as general as possible.

Each pixel of an image provides a linear constraint on the bi-variate BRDF by estimating the rendering equation. However the bi-variate BRDF representation induces a folding of the hemisphere since the light directions are symmetric about the view/normal plane in their BRDF domain which avoids the need to sample the full hemisphere or which generates 2 constraints in their 2D BRDF domain. They infer the BRDF from these constraints by creating an uniform grid and obtain a piecewise linear approximation of the BRDF. They note the noise caused by the sensor must be handled as well as the bivariate approximation, discretization of the rendering equation and error in the assumed shapes by adding a regularization term.

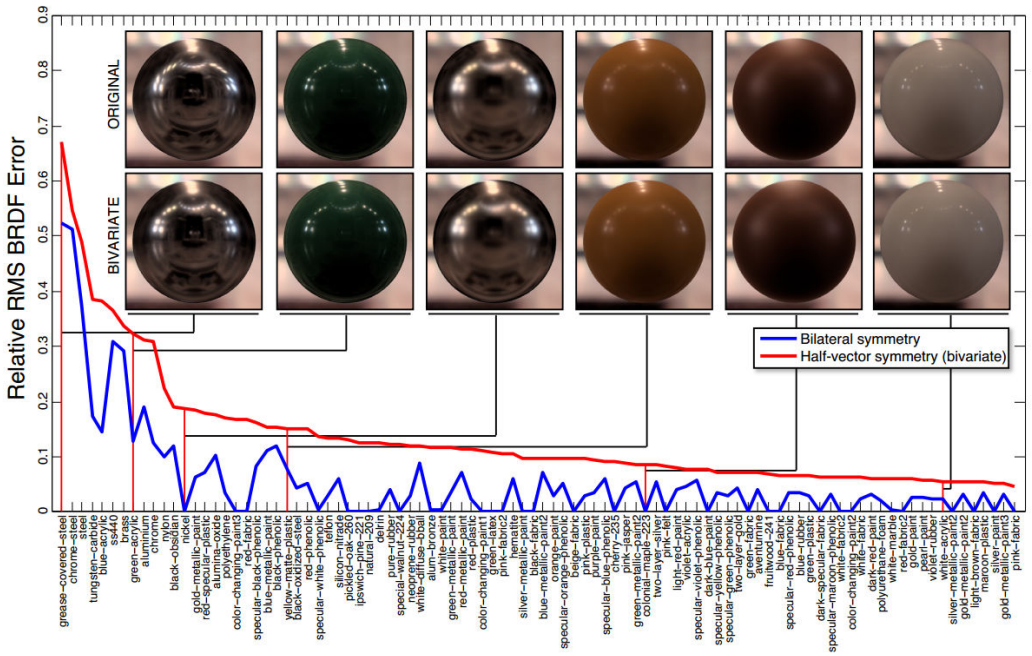


Figure 2.35: From [Romeiro *et al.*, 2008]. Red curve: Accuracy of bivariate representations of materials in the MERL/MIT BRDF database. Materials are in order of increasing accuracy, and while RMS BRDF error is seemingly large, rendered images reveal few perceivable differences. Blue curve: portion of RMS error explained by the original data's deviation from bilateral symmetry

.

2.9.2 Image based methods: unknown lighting

[Ramamoorthi and Hanrahan, 2001] introduce a signal-processing framework which describes the reflected light field as a convolution of lighting and BRDF under general illumination conditions. The inverse rendering problem can then be viewed as deconvolution by expressing the reflected light field as a product of spherical harmonic coefficients of the BRDF and the lighting. They demonstrate that in frequency space two important priors can be used in active reflectometry:

- the most appropriate object to recover lighting conditions is a highly reflective sphere since the frequency spectrum of a mirror BRDF is constant in this case.
- in frequency space recovering a filter using a delta function from its impulse response is a well-conditioned problem. In a consistent manner a directional light source is a delta function and a BRDF is a filter.

[Haber *et al.*, 2009] present an approach for recovering the reflectance of a static scene with known geometry from a collection of images taken under distant, unknown illumination. The main difference with previous approaches is the set of input views which include images taken under different illumination condition which allows them to recover reflectance even by using images from Flickr. Their method estimates incident illumination per-image at the same time as the surface point reflectance

Figure 2.36: Overview of [Haber *et al.*, 2009] reconstruction pipeline. From left to right: an example image taken from Flickr, re-rendered model using the recovered reflectance properties, the geometry and the illumination from the estimated environment.

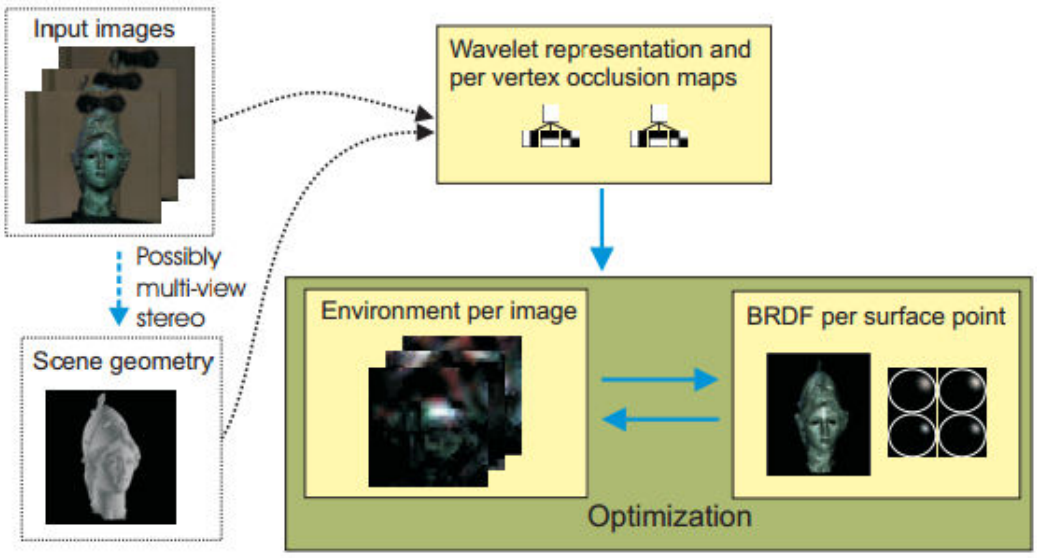


Figure 2.37: Overview of the system in [Haber *et al.*, 2009].

across views. Their method is based on a wavelet framework [Ramamoorthi and Hanrahan, 2001] to incorporate various reflection models. The interesting point is that they manage to estimate a BRDF per surface point and an illumination layer per input image through an iterative method. To do so, they constrain the BRDF estimation as a linear combination of basis BRDFs as presented by [Weistroffer *et al.*, 2007] thanks to the data acquisition of [Matusik *et al.*, 2003].

[Romeiro and Zickler, 2010a] propose to estimate BRDF by defining a prior probability distribution based on the image I for the unknown lighting L, $p(L)$ and the BRDF F, $p(F)$ and finding the functions that maximize $p(L, F|I) \alpha p(I|L, F)p(L)p(F)$. The notable feature of this approach is to se-

lect the BRDF that is the most likely under a distribution of light environments instead of selecting the single BRDF/lighting pairs that best explain an input image based on a probabilistic generative model. This Bayesian approach can then be combined with other techniques such as shape-from shading, contours, shadows as demonstrated in [Barron and Malik, 2013a] or [Chen and Koltun, 2013]. In [Romeiro and Zickler, 2010a] illumination is represented as spherical lighting using a Haar wavelet basis. Reflectance is restricted to be expressed as a linear combination of a positive basis function learnt through non-negative matrix factorization. During their optimization they consider also an exposure parameter to compensate for the difference between the absolute scale of intensity measurement and the scale between the illumination and reflectance functions learnt from normalized data. A scale ambiguity still exists for each image because increasing the overall brightness of the illumination will decrease the BRDF during the optimization. To avoid solving a color consistency problem by running their method for each channel, they perform inference on the luminance channel to recover a monochrome BRDF. The recovered red-material in Fig. 2.39 does not match the highlight colors of the reference image and cannot match it since the displayed input is the color image prior extracting luminance and the displayed output is the outer product of the recovered monochrome BRDF and the RGB vector. They evaluate their method using synthesized images with the MERL/MIT dataset of acquired BRDF of [Matusik *et al.*, 2003] and their own collection of measured HDR environment maps to setup ground truth comparisons. They still consider only a sphere with samples collected from 12 000 normals. They set up an experiment using synthetic input Fig. 2.38 and captured input Fig. 2.39; from these they predict the appearance of the material using the recovered BRDF and compare it to the ground truth. One important conclusion of the technical report [Romeiro and Zickler, 2010b] is that inferring explicit reflectometry in the wild may be possible to achieve.

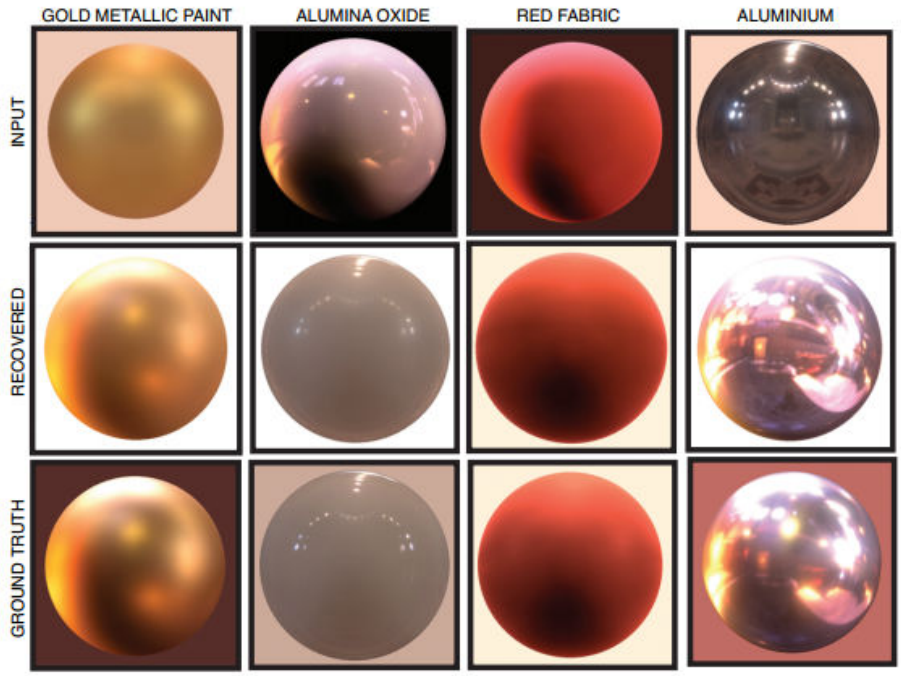


Figure 2.38: Evaluation with synthetic input [Romeiro and Zickler, 2010a]. Top: Single image used as input. Middle: Appearance predicted in a novel environment using the recovered BRDF. Bottom: Ground truth image in the same novel environment

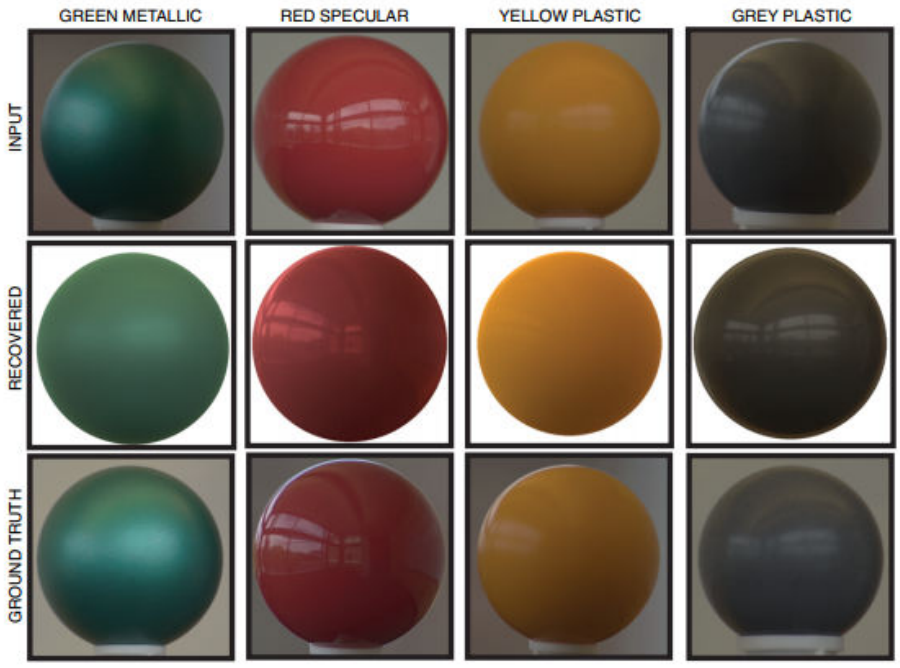


Figure 2.39: Evaluation with captured input [Romeiro and Zickler, 2010a]. Top: Image used as input. Middle: Appearance predicted in a novel environment using the recovered BRDF. Bottom: Ground truth images captured in the same novel environments

Chapter 3

Multi-View Intrinsic Images of Outdoors Scenes

In this chapter, we present our multi-view intrinsic image method. As outlined in Chapter 1, we target outdoors scenes with cast shadows, and wide-baseline datasets for easy capture. Our two key ideas are:

1) to progressively improve parameter accuracy with iterative estimation.

2) to express reflectance as a function of discrete visibility values, allowing the definition of a robust shadow classifier.

Our goal is to achieve high quality intrinsic image decompositions with as little user interaction as possible.

3.1 Image Model and Algorithm Overview

The image model we use is central to our method, since it clearly defines the quantities that need to be estimated. The model will also be used to guide the definition of our iterative process to estimate our multi-view intrinsic decomposition.

3.1.1 Image Model

We use the following image formation model [Laffont *et al.*, 2013]:

$$I \;=\; R\,(\; v_{\text{sun}}\, L_{\text{sun}}\; \cos(\omega_{\text{sun}}) + S_{\text{sky}} + S_{\text{ind}}), \tag{3.1}$$

I is the observed radiance (i.e., pixel value), R is the diffuse reflectance of the corresponding 3D point, L_{sun} is the radiance of the sun, v_{sun} is the sun visibility from the point, ω_{sun} is the angle between the normal n and the direction θ_{sun} to the sun, S_{sky} is the radiance of the visible portion of the sky

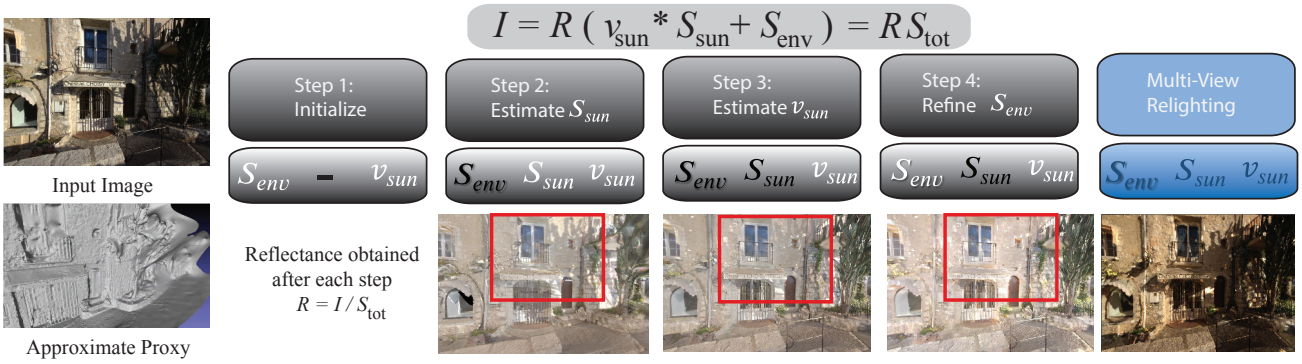


Figure 3.1: Our input are images (top left), an approximate 3D model (proxy) (lower left) and user-supplied sun direction. We use the image formation model (top) and estimate progressively better approximations to each of its parameters. In white, quantities estimated in a given step; quantities in black are fixed at that stage. Step 1: given the proxy we build a sky environment map and compute a first estimate of v_{sun} and S_{env} by ray-tracing the inaccurate 3D model and sky map. Step 2: we refine v_{sun} and estimate S_{sun} using luminance and chromaticity. Step 3: given first estimates of all quantities we perform a graph labelling to further refine v_{sun}. In Step 4 we refine S_{env}, and v_{sun} in penumbra, using the more accurate shadow boundaries now available. Reflectance is estimated at steps 2-4, and we clearly see how the result is progressively improved. Far right: during multi-view relighting, reflectance is fixed, and we can manipulate quantities in blue, resulting in a relit image (lower right; compare to top left).

integrated over the hemisphere Ω centered at n, and S_{ind} is the indirect irradiance integrated over Ω, but excluding the sky. For all cosines we take $\max(0, \cos)$ in practice; all values are RGB except for the cosine. We implicitly assume that R is diffuse.

Using Eq. 3.1, we can also write reflectance R as a function of visibility:

$$R(v_{\text{sun}}) = \frac{I}{(S_{\text{ind}} + S_{\text{sky}} + v_{\text{sun}}\ L_{\text{sun}}\ \cos(\omega_{\text{sun}}))} \tag{3.2}$$

In some cases it is convenient to group sky and indirect lighting into a single *environment shading* term S_{env} and write $S_{\text{sun}} = L_{\text{sun}}\ \cos(\omega_{\text{sun}})$, giving a simpler expression:

$$I = R\,(\,v_{\text{sun}} S_{\text{sun}} + S_{\text{env}}\,) \tag{3.3}$$

3.1.2 Input

Our input is a set of linearized raw 12-bit/channel photographs of the scene, captured from different viewpoints at the same time of day and with same exposure. We use Autodesk Recap360 (http://recap360.autodesk.com) for all 3D reconstructions, taking the vertices of the reconstructed mesh as a point cloud. The quality of the meshes is quite high overall with some residual noise for buildings, but often very approximate for structures such as vegetation. Such methods also have difficulty re-

constructing silhouettes and fine structures. Alternative methods (e.g., structure from motion [Snavely *et al.*, 2006] followed by reconstruction [Goesele *et al.*, 2007; Furukawa and Ponce, 2007; Pons *et al.*, 2007]) provide similar quality. In what follows we use the term *proxy* to refer to this – typically incomplete and inaccurate – 3D model.

Our method requires the sun direction θ_{sun}. Automatic methods [Panagopoulos *et al.*, 2013] can be used, however, we prefer to use a simple manual step, which is performed just after reconstruction and guarantees high-quality results. To determine the direction of the sun, a colored version of the point cloud is presented to the user. Each point is assigned the median value of pixels in all images in which this 3D point is visible. The user clicks on a point in shadow and the corresponding 3D point which casts it, allowing the sun direction to be estimated. This simple process is shown in the accompanying video.

3.1.3 Estimating image-model quantities

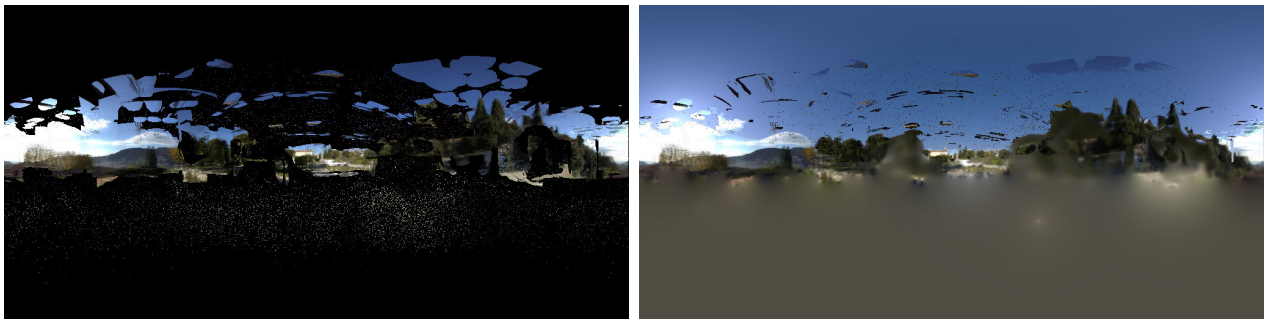
Our algorithm has four main steps, shown in Fig. 3.1. In each step we compute estimates of the quantities of Eq. 3.1, which are progressively more accurate. To compute a *reflectance* layer R, we estimate shading S_{tot}, and divide the input image to obtain R; this is performed in Steps 2-4 and the result shown in Fig. 3.1. In contrast with most previous work, our input contains strong cast shadows. Our goal is to obtain results of sufficient quality to perform relighting: this requires a reflectance layer free of shadow and other residues, as well a good estimate of shadow boundaries, environment shading.

A guiding principle of our approach is that we prefer the explanation of a given scene that favors a smaller number of reflectances, following previous work [Omer and Werman, 2004; Barron and Malik, 2013b; Laffont *et al.*, 2013]. Consider the scene shown in the inset. There are two explanations for the dark areas on tablecloth: a shadow cast by the statue and plant or blobs painted in gray. Our approach favors the hypothesis with fewer reflectances, which explains the image as a shadow over a uniformly white tablecloth. Throughout the four steps of our approach, we enforce this hypothesis by finding same-reflectance *pairs* between regions or points in light and shadow, inspired by previous work [Panagopoulos *et al.*, 2013; Guo *et al.*, 2011].

The key novelties of our approach are the automatic estimation of the parameters of Eq. 3.2, and the introduction of a robust shadow classifier using this information, see Sec. 3.4. Put together, these encourage the choice of the correct visibility configuration which finds same reflectance regions and implicitly connects (or merges) them via the pairs, thus enforcing the hypothesis.

The four steps are illustrated in Fig. 3.1: In Step 1, we find initial values for S_{env} and v_{sun}; in Step 2

we estimate S_{sun}, in Step 3 we obtain accurate shadow boundaries by refining the estimate of v_{sun} and in Step 4 we refine the estimation of S_{env}. As we can see in the figure 3.1, the resulting reflectance at each step is significantly improved.



(a) (b)

Figure 3.2: Left: the partial environment map (a). Right: completed synthesized environment map (b).

(a) (b) (c) (d) (e)

Figure 3.3: The consecutive steps of the algorithm to determine L_{sun} for the "Street" scene. (a) Input image (b) shadow from inaccurate 3D model: the proxy overestimates the geometry of the cactus and creates a "blob" shadow (c) K-means intensity estimation: some black areas are not shadows (d) intersection of (b) and (c): a more reliable subset of shadows are found, which are used in (e) to find pairs used to calibrate the sun.

3.2 Initialization: Estimation of S_{sky} and S_{ind}

To compute S_{sky} and S_{ind}, we first automatically compute an environment map to represent light coming from the sky and unreconstructed surfaces[1]. We project all pixels of the input pictures that are not covered by the reconstructed geometry into this map. Fig. 3.2 shows such a partial environment map where holes correspond to directions either not captured in the input photographs or directions corresponding to rays that do not intersect the proxy.

We apply a simple color-based sky detector to determine which pixels above the horizon in the map are sky and which are distant objects. More involved approaches [Tao *et al.*, 2009] could be used, but our approach sufficed in all our examples. The horizon is the main horizontal plane of the proxy. The visible portions of the sky give us strong indications on the atmospheric conditions at the time

[1] We described a preliminary version of the environment map computation in Chapter 4 of [Laffont, 2012].

of capture. Inspired by Lalonde *et al.* [2009, 2012], we estimate the missing sky pixels by fitting the parametric sky model of Perez et al. Perez *et al.* [1993] from the partial environment map. This model expresses for any direction p the sky color *relative* to the color at zenith as a function of the angle θ_p between p and the zenith, the angle γ_p between p and the sun direction, and the turbidity t that varies with weather conditions [Preetham *et al.*, 1999; Lalonde *et al.*, 2009]. Since the color at zenith is itself an unknown, we need to recover a global per-channel scaling factor to obtain absolute values. We estimate the turbidity t of the sky model f and the scaling factor $\mathbf{k}$ by minimizing

$$\underset{t,\mathbf{k}}{\operatorname{argmin}} \sum_{p \in \mathcal{P}} (\mathbf{k} f(\theta_p, \gamma_p, t) - \mathbf{A}_p) \tag{3.4}$$

where $\mathcal{P}$ denotes the set of known pixels in the environment map $\mathbf{A}$. We solve this non-linear optimization with the simplex search algorithm (`fminsearch` in Matlab). At each iteration, the search algorithm generates a new value of the turbidity t that we use to update f, and then $\mathbf{k}$ from the new sky values by solving a linear system. We initialize the optimization by setting $t = 3.5$, which corresponds to the turbidity of a clear sky [Preetham *et al.*, 1999]. We fill holes below the horizon line by diffusing color from nearby pixels.

Similarly to Laffont *et al.* [2013], we compute S_{ind} and S_{sky} by integrating the indirect and sky incoming radiance using ray-tracing. For each 3D point, we cast a set of rays over the hemisphere centered on the point normal. Rays that intersect the sky part of the environment map contribute to S_{sky}, while rays that intersect the proxy geometry or the non-sky pixels of the environment map contribute to S_{ind}. We estimate the radiance coming from the proxy geometry by gathering for each vertex the radiance in the images where this vertex appears. We assign the median of the gathered values as the approximate diffuse radiance of the vertex. Given the low frequency nature of these quantities, our approximations are generally sufficient. However, the non-diffuse nature of real surfaces and errors in reconstruction can result in overestimation of indirect light. We thus introduce an approximate attenuation factor which compensates for such errors by scaling with the cosine of the normal of the contributing surface when gathering at each point.

Compensating for Superfluous Indirect Light The outdoors scenes we target contain perpendicular and horizontal surfaces (walls, floors, etc.). The reconstruction of such corners is often incorrect, with geometry being added to the proxy. We often observe such geometry at grazing angles in the photographs, resulting in a high median value. When gathering indirect light at a given point x this can result in a higher contribution from such points. Finding the correct attenuation factor would require complete geometry and BRDF data, so we can only provide an approximate scale factor. Consider such a point x at which we gather light, and a point y on another surface contributing to x. The incoming angle θ_i is the angle between the direction $y - x$ and the normal n_y at y. We attenuate incoming lighting

by $\cos\theta_i$, thus reducing the contribution at grazing angles, which is amplified by the incorrect reconstruction. This is a coarse approximation, but is well adapted to the case of perpendicular surfaces such as walls and ground which are predominant in outdoor scenes. This approach improves the result in all scenes we tested, in particular in regions containing evidently non-diffuse surfaces.

The ray-tracing step also provides approximate visibility $\tilde{v}$ towards the sun at each point, with respect to the proxy. The boundaries defined by $\tilde{v}$ can be quite approximate however, as shown in Fig. 3.3(b). We improve the estimate of v_{sun} in Step 3 (Sec. 3.4).

3.3 Estimation of Sun Color L_{sun}

Now that we have computed illumination from the sky and indirect transfer at all 3D points, we can estimate L_{sun} using Eq 3.1 and a pair of points with same reflectance and different visibility. Given two points p_1 and p_2 with the same reflectance, with one in shadow and the other in light, we can compute L_{sun}:

$$L_{\text{sun}} = \frac{I_1 * (S_{\text{sky2}} + S_{\text{ind2}}) - I_2 * (S_{\text{sky1}} + S_{\text{ind1}})}{I_2 * v_{\text{sun1}} * \cos(\omega_1) - I_1 * v_{\text{sun2}} * \cos(\omega_2)} \tag{3.5}$$

All quantities for sun, sky and indirect are denoted with appropriate subscripts.

The main difficulty in using this formulation is that we do not yet have accurate reflectance and visibility necessary to find a suitable pair of points. While single-image intrinsic decomposition methods could be used to initialize the reflectance, most existing algorithms are challenged by outdoor scenes with hard shadows that break the Retinex assumptions of a smooth monochrome shading. We conducted preliminary experiments with the Retinex implementation of [Grosse *et al.*, 2009], which confirmed that this algorithm does not remove hard shadows on our scenes. As a result, our calibration algorithm was unable to find pairs of points sharing the same reflectance across shadow boundaries.

Instead of using Retinex, we found it sufficient at this stage to approximate reflectance with the image chrominance and shading with luminance, which we combine with the proxy-based visibility $\tilde{v}$ for a conservative estimate of shadow regions. More precisely, we first perform a K-means clustering on luminance (with $K = 4$), and classify a cluster in shadow if its ratio of points inside and outside the proxy shadow $\tilde{v}$ is lower than the average ratio of the K clusters. We then intersect the value of $\tilde{v}$ (Fig. 3.3(b)) with the classification of each pixel (Fig. 3.3(c)), resulting in very confident regions of shadow albeit covering only a limited number of pixels in the image Fig. 3.3(d). Given this visibility estimate, we sample the shadow boundary regularly and for each sample we detect a lit (resp. shadowed) point away from the penumbra by walking in the two directions perpendicular to the boundary and selecting the pixel with the highest (resp. lowest) luminance and a similar chrominance. In our implementation we stop the walk after 30 pixels in each direction and reject the samples for which no

pixels with similar chrominance are found. We also reject the sample if we cross a depth or normal discontinuity along the walk, identified with a Canny filter over the depth and normal map of the proxy. Taken over the entire multi-view dataset, these pairs provide multiple estimators for L_{sun}, using Eq. 3.5 (Fig. 3.3(e)).

We finally compute a robust estimate of L_{sun} as the median of the solutions given by all pairs.

We found that performing the median filter in each RGB channel separately gives the best results. This sparse set of pairs is approximate but sufficient for the calibration task. The later estimation of more accurate visibility boundaries will allow us to find a more reliable and denser set of light/shadow pairs and thus refine the estimate of environment lighting.

At this stage, we have an estimate of all quantities of Eq. 3.1, namely ω_{sun}, S_{sky}, S_{ind} and L_{sun}; the estimate of $v_{\text{sun}} \approx \tilde{v}$ however is approximate. If we compute reflectance at each 3D point, we obtain approximate results that can have large regions of error (see leftmost image in Fig. 3.1).

3.4 Estimating Accurate Cast Shadows and Intrinsic Layers

To compute residue-free reflectance layers for each image, we need to refine the accuracy of shadow boundaries and thus v_{sun}. We do this using a graph labeling approach, giving a binary label of shadow or light to all pixels, except those in penumbra. We assign a continuous visibility label to penumbra separately using matting (Sec. 3.4.2).

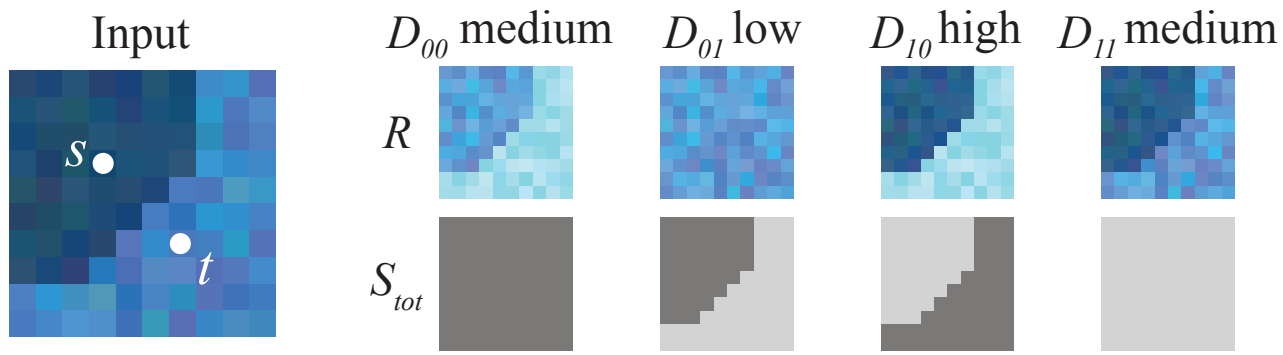
3.4.1 Shadow Labeling

The intuition behind our approach is to find the set of visibility labels that make most points share a similar reflectance, as explained earlier (Sec. 3.1.3). Consider two points s and t with visibility i and j respectively. Using Eq. 3.2 we compute the difference between their reflectances as:

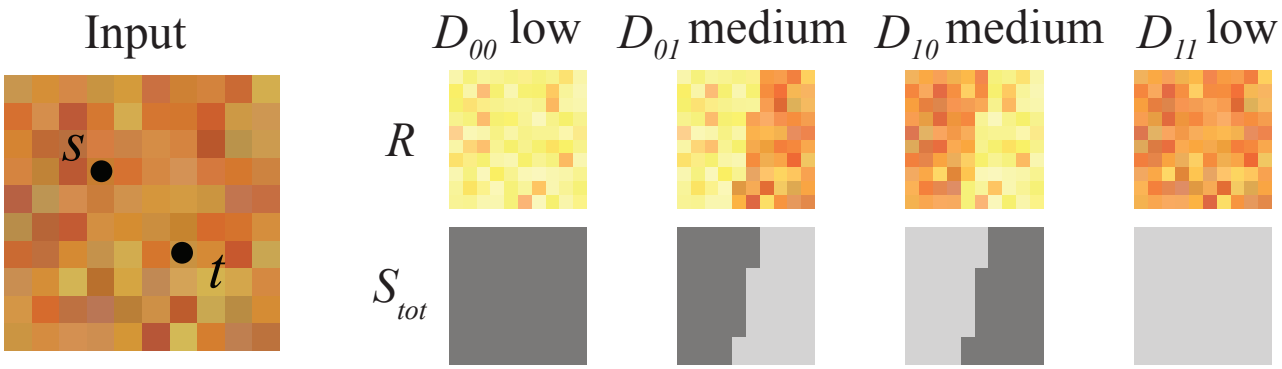
$$D_{ij} = |R_s(i) - R_t(j)|. \tag{3.6}$$

Since $i, j \in \{0, 1\}$ we obtain four possible values of D_{ij}. A small value provides us with a strong evidence that s and t share the same reflectance under the corresponding visibility hypothesis. We illustrate this strategy with the following toy examples:

The diagram below shows the case where s is in shadow and t is in light, both on a patch of roughly constant reflectance. Consider the case in the second column, which is the correct configuration: the two points receive a similar reflectance, which makes D_{01} small. In contrast, D_{10} is large, since the incorrect visibility assumptions "pull apart" $R_s(1)$ from $R_t(0)$; see Eq. 3.2. The two points also receive different reflectances when assigned the same visibility, i.e. D_{00} and D_{11} are larger than D_{01}, although



smaller than D_{10}. We can thus concentrate on comparing D_{10} and D_{01}; this provides a robust indicator of the correct visibility labels for the pair, under the assumption that the two points share a similar reflectance. Importantly, this information is directional, i.e., if s is in shadow, then we have a strong indication that t is in light.



The second diagram above shows a configuration where s and t have the same label (both in light in this case; both in shadow can be treated symmetrically), also with the same reflectance. Here we can distinguish clearly between same label cases (D_{00} and D_{11}) which give a similar reflectance to s and t compared to the different-label cases. However, we cannot distinguish between the light/light or shadow/shadow case since they both make the two points have a similar reflectance. Pairs of points sharing the same visibility are thus somewhat less informative than pairs of points with different visibility. Both cases however provide reliable information which we use for shadow classification. Finally, points having different reflectances result in high D_{ij} under all four labeling configurations.

We next define an energy that is minimized by the label configuration best explaining the same reflectance hypotheses. Specifically, we detect the pairs of points likely to have the same reflectance and different visibility and use this directional information to initialize the labels at a few confident points (Fig. 3.4(a)). We then connect these points to their immediate neighbors and to other points with same reflectance and visibility, which allows us to propagate the labeling over the entire image (Fig. 3.4(b)). We express this approach as a Markov Random Field (MRF) problem over a graph [Szeliski, 2010a; Kolmogorov, 2006], where each node corresponds to a point s with label $x_s \in \{0, 1\}$ and each edge

(s,t) connects a point s to another point t. Noting $\mathcal{X}$ the set of all labels x_i of all nodes, we have

$$\underset{\mathcal{X}}{\operatorname{argmin}} \sum_{s \in V} \phi_s(x_s) \; + \; \sum_{(s,t) \in \mathcal{E}} \phi_{s,t}(x_s, x_t), \;\; x_i \in \{0,1\}. \tag{3.7}$$

V denotes the set of nodes, $\mathcal{E}$ is the set of edges, $\phi_s(x_s)$ is the unary potential deduced from points with same reflectance and different visibility, and $\phi_{s,t}$ is the pairwise potential that favors the propagation of the labels. We detail the computation of the unary and pairwise potentials later in this section.

(a) Initialization (b) Final labeling

Figure 3.4: (a) Initial labels from unary term, white is in light, black in shadow and grey undefined. (b) Final labels after convergence.

To solve this optimization, we could naively connect all pixels to all others, and perform the minimization on the resulting graph. This is both inefficient and numerically unstable. We thus apply mean-shift clustering [Comaniciu and Meer, 2002] in (L,a,b,x,y) space to segment the image into small regions where we can safely assume uniform reflectance and visibility, simplifying the problem and reducing noise (see Fig. 3.5). The values for $R(0)$ and $R(1)$ for a cluster are computed as the median values for all 3D points projected onto the cluster, except for points in a 3-pixel wide boundary around each cluster.

We solve our problem using a publicly available implementation of [Kolmogorov, 2006][2]. The potentials take values $l_1 = 1$, $l_0 = 0$ when strongly encouraging one hypothesis over the other, $l_p = 0.8, l_{np} = 0.2$ for the case when one hypothesis is moderately preferred over another ("non-preferred") and $l_{eq} = 0.5$ when both hypothesis are equally encouraged.

Unary Potential. As many binary labeling problems, a good initialization is central to obtain a good solution. Given the discussion above, we use pairs of points likely to have the same reflectance

[2]http://research.microsoft.com/en-us/downloads/dad6c31e-2c04-471f-b724-ded18bf70fe3/

(a) Input (b) Meanshift clustering (c) Cluster boundaries

Figure 3.5: We apply meanshift clustering to decompose the image in small regions of uniform color. We then solve the shadow labeling on a graph of clusters rather than pixels, which reduces the number of unknowns and the impact of noise.

and different visibility to initialize our unary term. In particular, for a cluster s we find the set $\mathcal{S}$ of k other clusters with the smallest D_{01} and the set $\mathcal{L}$ of k other clusters with the smallest D_{10}. The clusters in $\mathcal{S}$ favor the hypothesis that s is in shadow, while the clusters in $\mathcal{L}$ consider that s is in light. We compute the score of each hypothesis as the sum of the reflectance differences between s and the k other clusters

$$H_0 = \sum_{t \in \mathcal{S}} |R_s(0) - R_t(1)| \tag{3.8}$$

$$H_1 = \sum_{t \in \mathcal{L}} |R_s(1) - R_t(0)| \tag{3.9}$$

If $\frac{H_0}{H_1 + H_0} < \tau_1$, i.e., the hypothesis that s is in shadow is stronger, we set the unary potential to prefer the "in shadow" label:

$$\phi_s(x_s) = \begin{cases} l_0 \text{ for } x_s = 1 \\ l_1 \text{ for } x_s = 0 \end{cases} \tag{3.10}$$

Conversely, if $\frac{H_1}{H_1 + H_0} < \tau_1$, we set the unary potential to prefer the label "in light":

$$\phi_s(x_s) = \begin{cases} l_1 \text{ for } x_s = 1 \\ l_0 \text{ for } x_s = 0 \end{cases} \tag{3.11}$$

If neither condition is true we perform a more localized search. We compute two new hypothesis

H_1' and H_0' in the same manner as Eq. 3.8, but restrict the k clusters to lie within a *neighborhood* around s. We then check if:

$$\frac{H_1 \ + \ H_1'}{H_0 \ + \ H_1 \ + \ H_0' + H_1'} < \tau_1 \tag{3.12}$$

and similarly for the H_0 hypothesis, which can be seen as a more "permissive" hypothesis, since we complement the best global candidates with the best local ones. If one of these conditions is met, we set the potentials the same way as above. If none of the conditions are met, the unary potentials are set to equally prefer either hypothesis:

$$\phi_s(x_s) \ = \ l_{eq}, \quad x_s \in \{0, 1\} \tag{3.13}$$

We used $\tau_1 \ = \ 0.1$, corresponding to a 90% confidence level required to make a decision.

Pairwise Interaction Potentials. The goal of our pairwise potentials is to propagate labels between clusters with the same visibility. We first create edges between each cluster s and other clusters with similar reflectance, which we select as the k clusters with smallest D_{00} or D_{11}. For these edges, the values of the potentials are set to strongly encourage the same label to be propagated:

$$\phi_{s,t}(x_s, x_t) \ = \ \begin{cases} l_1 \ \text{when} \ x_s \ = \ x_t \\ l_0 \ \text{when} \ x_s \ \neq \ x_t \end{cases} \tag{3.14}$$

However, these edges alone are not always sufficient to ensure that the graph forms a single connected component. We prevent isolated components by also connecting each cluster with its immediate neighbors. In the absence of other cues, we define the potential of these weaker connections to encourage clusters with the same color distribution to share the same visibility. We compute the χ^2 histogram distance d_c in Lab space for clusters s and t using the approach described in [Chaurasia *et al.*, 2013]. Clusters s and t are similar for $d_c \ < \ \tau_c$; in this case we assume they most probably have the same label:

$$\phi_{s,t}(x_s, x_t) \ = \ \begin{cases} l_{np} \ \text{when} \ x_s \ = \ x_t \\ l_p \quad \text{when} \ x_s \ \neq \ x_t \end{cases} \tag{3.15}$$

If the χ^2 distance is too large however, all potentials are set to equally prefer all possible hypotheses.

$$\phi_{s,t}(x_s, x_t) \ = \ l_{eq}, \ \ x_s, x_t \in \{0, 1\} \tag{3.16}$$

We used $\tau_c \ = \ 0.05$ for all our tests, which corresponds to the acceptance probability in the χ^2 test.

At convergence, we obtain accurate shadow boundaries, even though there can be some occasional

miss-classifications, e.g., the letters on the store front in Fig. 3.4(b). Such errors typically occur in small regions that contain few or no 3D points. In the former case, the median reflectance candidates $R(0)$ and $R(1)$ are more likely to be polluted by occasional reprojection errors and specularities, while in the latter case the propagation is solely governed by the χ^2 distance to neighboring regions. Nevertheless, erroneous regions tend to be small in size , and thus do not affect the application to relighting.

3.4.2 Per-pixel Estimation of v_{sun} and Intrinsic Layers

The binary labeling cannot capture soft shadows. We apply Laplacian matting [Levin *et al.*, 2008a] to recover continuous variations of visibility in the boundaries between clusters. These correspond to penumbra regions at the frontier of shadow and light clusters, effectively providing a tri-map from the binary shadow mask. We also apply Laplacian matting guided by the input image to propagate the shading values $S_{\text{sky}},\ S_{\text{ind}}$ and S_{sun}, as previously done by Laffont et al. Laffont *et al.* [2012]. We use all 3D points except those in the boundaries between clusters as constraints in this propagation. While these smooth shading layers do not contain shadows, propagating them using the input image as guidance sometimes produces artifacts along shadow boundaries. We reduce these artifacts by excluding a small band along shadow boundaries from the propagation, which we subsequently fill with a color diffusion. The reflectance layer is obtained by dividing the input image by the sum, or total shading S_{tot}

$$S_{\text{tot}} \;=\; S_{\text{sky}} \;+\; S_{\text{ind}} \;+\; v_{\text{sun}} S_{\text{sun}}. \tag{3.17}$$

The classifier can occasionally miss very fine shadow structures which are however captured by the clusters; we also propagate visibility in the boundary regions between clusters, which generally improves the visual quality for relighting (see Chapter. 4).

3.5 Refining Environment Shading and Reflectance Estimation

3.5.1 Our approach

The quality of the intrinsic layers obtained so far is limited by the accuracy of the different radiometric quantities computed. In particular, the success of using Eq. 3.2 to compute R is dependent on the approximations in our estimation of S_{env}, L_{sun} and v_{sun}. As we see in Fig. 3.6(a), the currently estimated values leave a visible residue in the reflectance layer, which should be continuous (Fig. 3.6(c)). This discontinuity occurs because the values of S_{env} and L_{sun} were computed using the incomplete and inaccurate 3D reconstruction, and are thus approximate.

We illustrate this in Fig. 3.7 where we show a plot of image intensity across a shadow boundary,

(a) Reflectance before correction (b) Same-reflectance pairs across shadow boundary (c) Reflectance after correction

Figure 3.6: (a) The reflectance is discontinuous across the shadow boundary due to incorrect estimation of shading. (b) Pairs chosen as constraints to impose the same reflectance on both sides of the boundaries. (c) Corrected reflectance after optimization.

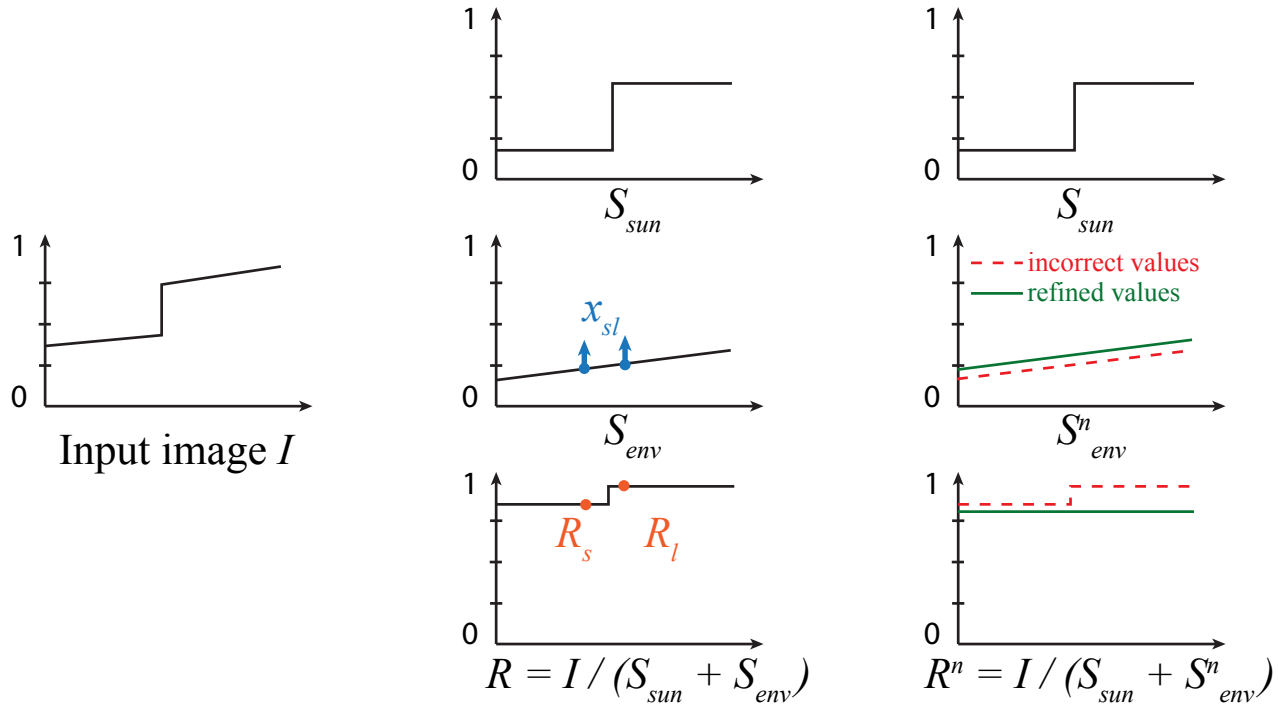


Figure 3.7: 1D visualization of S_{env} refinement. Small errors in our estimates of L_{sun} and S_{env} can prevent the reflectance to be continuous across shadow boundaries (middle). We detect pairs of points with similar reflectance on each side of the boundary (orange dots) and compute a local offset of S_{env} (blue) that makes the two reflectances equal (right).

with the shadow region on the left. In the middle column, we see the decomposition of the image into reflectance (R_s in shadow and R_l in light), and shading, composed of S_{sun} and S_{env}. We will refine the value of shading so that R_s becomes equal to R_l, by adding an offset to S_{env}. We correct S_{env} since it

is a continuous quantity over the shadow boundary. Specifically we apply an offset x_{sl} to S_{env} on both sides of the shadow boundary so that R_s becomes equal to R_l (Fig. 3.7, right).

We first find a dense set of same reflectance light/shadow pixel pairs along the shadow boundaries, Fig. 3.6(b). For each pair, we compute an offset x_{sl} which makes the two reflectances equal. We then smoothly propagate the offsets to all pixels while preserving the variations of S_{env}, yielding the refined layer S^n_{env}, Fig. 3.6(c). The values of v_{sun} in penumbra were determined by image-driven propagation, which can sometimes result in high-frequency inaccuracies of v_{sun}. These cannot be captured by the smooth propagation, and we thus treat these pixels separately by correcting the v_{sun} layer.

Implementation details of the above steps for S_{env} refinement are described in the following subsection.

3.5.2 Implementation Details of S_{env} Refinement

To refine the estimation of S_{env} we first find a set of light/shadow pairs, we then compute the offset values x_{sl} and propagate the refined S^n_{env} values over the image. The implementation has two main steps: finding pairs and offset values and smooth propagation.

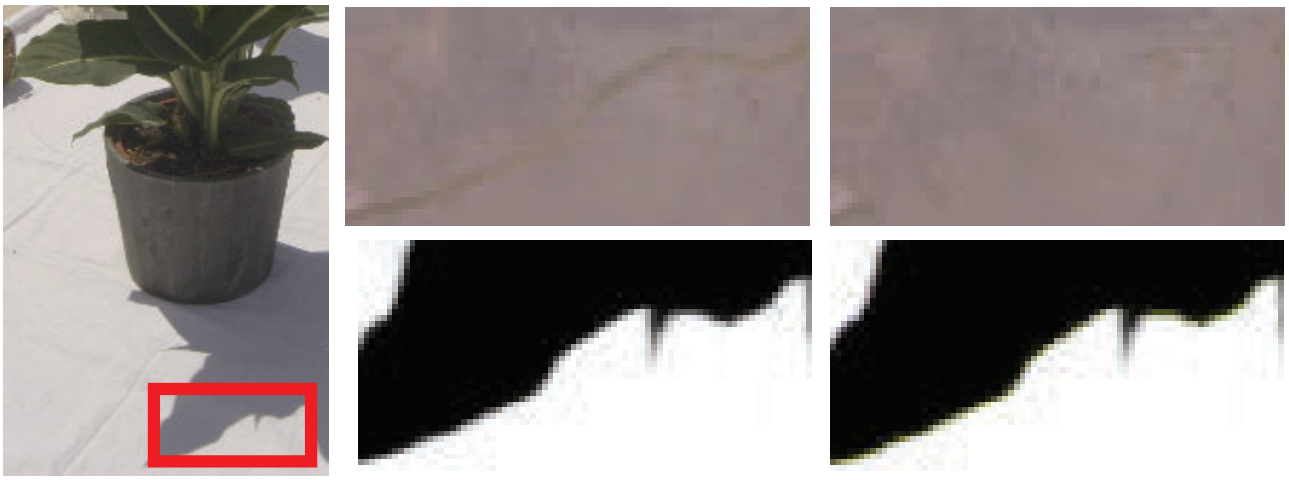


Figure 3.8: The reflectance contains halo artifacts in penumbra regions due to errors in the visibility (top middle). We re-estimate the visibility (bottom, mid and right) to remove these artifacts (top right). The differences in visibility are very subtle, please zoom into the pdf to see them.

Pairs and Offset Values. We find pairs by traversing shadow boundaries, in a manner similar to the L_{sun} estimation process (Sec. 3.3). We keep pairs with same reflectance, which we identify by a small D_{ij} value, since the visibility labels i and j are mostly correct. We also only keep pairs that satisfy the chromatic alignment of shadow/light pairs used in [Guo *et al.*, 2011]; we thus avoid creating pairs on incorrectly classified boundaries.

For each pair, we add an offset x_{sl} to S_{env} to make the two reflectances equal:

$$R_s = R_l \Rightarrow \frac{I_s}{v^s_{\text{sun}} S^s_{\text{sun}} + S^s_{\text{env}} + x_{sl}} = \frac{I_l}{v^l_{\text{sun}} S^l_{\text{sun}} + S^l_{\text{env}} + x_{sl}} \tag{3.18}$$

Re-arranging the terms gives the offset value:

$$x_{sl} = \frac{I_s(v^l_{\text{sun}} S^l_{\text{sun}} + S^l_{\text{env}}) - I_l(v^s_{\text{sun}} S^s_{\text{sun}} - S^s_{\text{env}})}{I_l - I_s} \tag{3.19}$$

Smooth propagation. The pairs of light/shadow pixels provide us with the values of $S^n_{\text{env}} = S_{\text{env}} + x_{sl}$ along the shadow boundaries. We propagate this information to all pixels by solving for the S^n_{env} image that minimizes

$$\underset{S^n_{\text{env}}}{\text{argmin}} \sum_{\partial S} ||S_{\text{env}} + x_{sl} - S^n_{\text{env}}||^2 + \sum_{\mathcal{P}} ||\nabla S_{\text{env}} - \nabla S^n_{\text{env}}||^2 \qquad + w \sum_{\mathcal{P}} ||S_{\text{env}} - S^n_{\text{env}}||^2$$

where ∂S is the set of constrained pixels along the shadow boundaries and $\mathcal{P}$ is the set of all image pixels. The first term encourages the constraint satisfaction, the second term preserves the variations of the original S_{env}, and the last term is a weak regularization that encourages the solution to remain close to S_{env} away from the shadow boundaries, using a small weight $w = 0.01$. This optimization can be solved using any standard least squares solver (we use the `backslash` operator in matlab).

Since x_{sl} can be negative, we can obtain negative values of S^n_{env} for a very small number of pixels. This can occur for example in regions which are poorly reconstructed as cavities, resulting in S_{env} values close to zero. We iterate by adding constraints for such points, setting $x_{sl} = 0$ such that S^n_{env} is equal to S_{env}. In all our experiments a single iteration was required to remove all negative values, which were always less than 1% of the pixels in the image.

Correcting Penumbra. The re-estimation of S_{env} described above ensures that both sides of a hard shadow boundary receive the same reflectance. However, errors also occur in the penumbra regions due to approximate continuous visibility, yielding halo artifacts in these regions (Fig. 3.8(mid top)). We correct these visibility values by associating each penumbra pixel to its closest pair of same reflectance light/shadow pixels as detected above. We then deduce the value of v_{sun} that makes the pixel receive the same reflectance. Fig. 3.8(right top) shows the final corrected reflectance. The effects are overall quite subtle, but this step does improve the result overall.

3.6 Intrinsic Decomposition Results

We present results on a variety of scenes. We show two test scenes with a small number of objects (Plant, Fig. 5.12) and Toys (Fig. 3.9, top row). We also show three natural scenes with buildings, vegetation and thin structures (Fig. 3.9). In most cases we obtain reflectance layers with little shadow and lighting residue, which are thus suitable for relighting. The shadow classifier and visibility layers are also of high quality overall; occasional miss-classifications are usually in small regions, which can be detected and be removed when moving the shadows for relighting. The strongest errors occur in scenes with poor geometric reconstruction, as is the case in the second and third row of Fig. 3.9 where large portions of the tree as well as the small wall in front of the scene are missing. Such holes in the geometry affect all the steps of our algorithm, from the computation of indirect lighting to the initialization of shadow regions and sun calibration. As a result, our shadow classifier has moderate success in identifying the shadow over the ground. Finally, the ground is dominated by variations of grey reflectance, which adds to the difficulty of shadow detection as some of these variations are well explained as shadows. The extended results for datasets are provided in section 5.4.1.

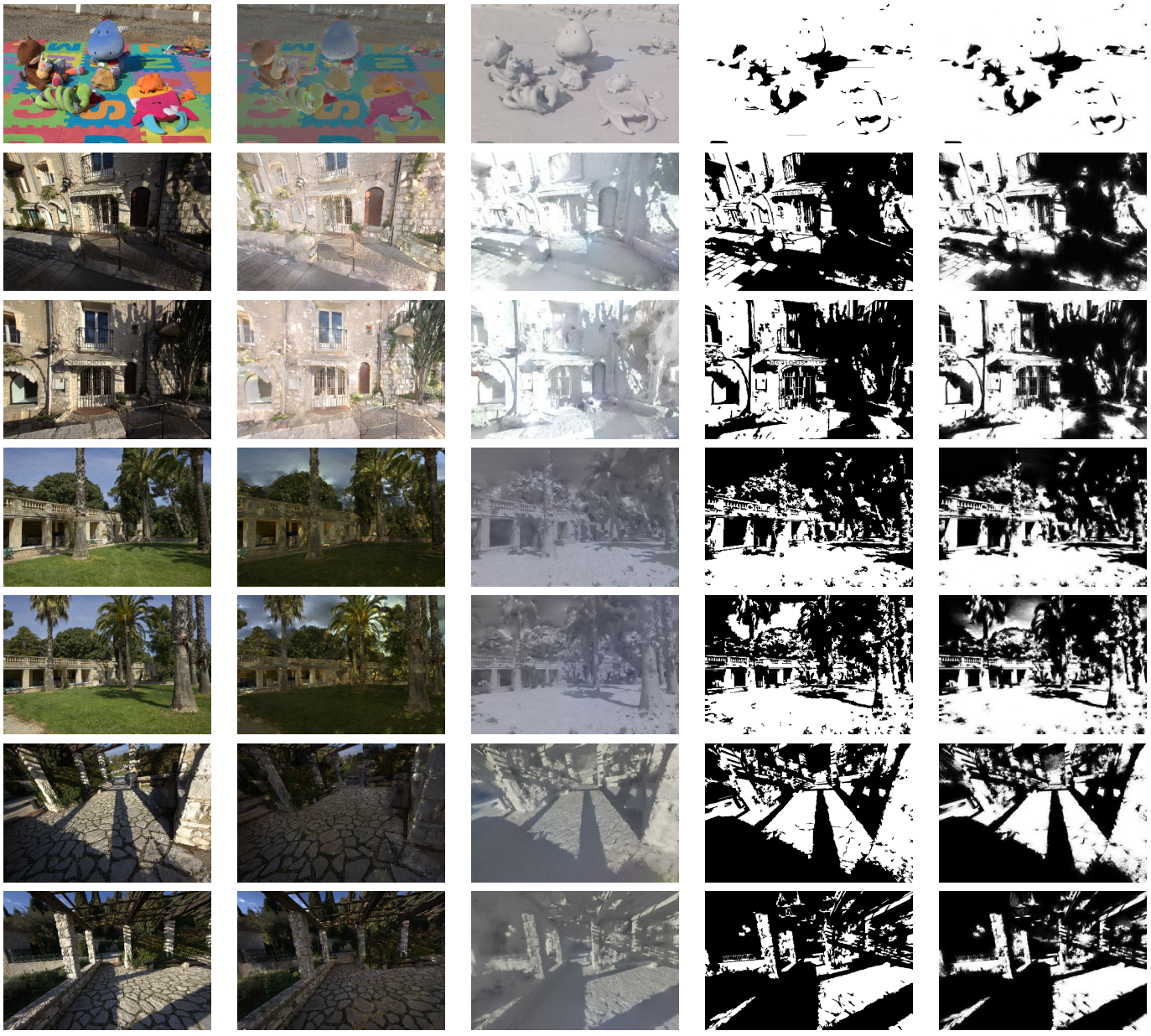


Figure 3.9: Our extracted layers on a variety of scenes: toys, urban (top), vegetation (middle), thin structures (bottom).

3.7 Conclusion

In conclusion, we see that our progressive estimation approach, together with our robust shadow classifier allow us to generate high-quality intrinsic image decompositions for multi-view datasets. As we shall see in the following chapter, these decompositions are of sufficient quality to achieve relighting of multi-view datasets, which can be used e.g., for image-based rendering (IBR) with changing lighting conditions. In Chapter 5, we present more results and extensive evaluation of our intrinsic image algorithm, while in Chapter 6 we will discuss initial ideas and experiments on going beyond the diffuse reflectance assumption.

Chapter 4

Relighting algorithms for multi-view image datasets

Image based methods are limited by the lighting conditions of the capture. This is an inherent problem which concerns image-based rendering applications such as Google street view, VFX compositing, matte painting. With tools such like Photosynth[1], 123DCatch [2] anyone can reconstruct their house, garden or a monument at one particular time of the day; but changing the time of day in the image is currently very hard.

This chapter describes a new relighting algorithm which can be used with our intrinsic decomposition pipeline for outdoor scenes described in chapter 3. We have an image collection of a scene, its 3D reconstruction and for each input image its decomposition into a collection of layers: reflectance, sky irradiance, indirect illumination, sun visibility, sun shading, normal and a residual layer to re-balance misestimation. By analogy with VFX practices, in most cases the best way to manipulate a scene is to work with easily editable layers. The most common task for artists is to compose an action shot captured in a studio or in a green stage of an environment shot (CG or real). Environments are getting more and more complex, and artists mix several part of shots, 3D reconstructions and full models to create them. Unfortunately, it is almost impossible to capture all these pieces to get coherent lighting and it requires a lot of work involving highly-skilled artists to fix these environments by editing each part separately. HDR probes are used to capture the environment lighting to ensure that artists can easily reproduce lighting conditions including plausible reflections.

The seven minutes ”barrel” sequence from the movie ”The Desolation of Smaug” required 98 hours of footage from aerial shots, green-screen sets, live-action shots, complex CG environments as well as having to build a swimming pool in a studio shown in Fig. 4.1. This scene has complex

[1] Windows solution for reconstruction.

[2] Autodesk product dedicated to multi view reconstruction.

Figure 4.1: Real footage courtesy Weta from wired.com.

camera paths to track with the decor during the shot, then such paths must be reproduced during capture. Challenges occur when realigning studio shots with the environment especially when lighting conditions and texture are very different. It is probably one of the most expensive scenes over the last decade.

In the case of Google street view one recurring problem is due to the method of acquisition itself. Google cars cannot capture all streets at the same time and do not have a consistent lighting. Our relighting method can offer solutions to both of these problems.

The most common application for intrinsic decomposition focuses on editing the reflectance texture to change the appearance of an object. Since our images are decomposed into reflectance and illumination layers, we can easily perform this modification.

Some work has been done on environment lighting estimation to allow quick insertion of CG elements or for augmented reality application [Lalonde *et al.*, 2012], [Karsch *et al.*, 2014b]. For lighting, we have our illumination layer split in indirect, sky and sun shading and also an environment map, while our method we can also insert an object easily as shown in Fig. 4.19.

All of these previous approaches do not support strong lighting condition changes such as moving the sun. With the development of automatic 3D reconstruction solutions (**PMVS**, **123D Catch**,...) and image based rendering methods [Buehler *et al.*, 2001], [Eisemann *et al.*, 2008] [Chaurasia *et al.*, 2011], [Chaurasia *et al.*, 2013], [Lipski *et al.*, 2014] being limited to the captured lighting condition of a scene is not only a problem for FX industries, but for all applications which require the rendering of captured

Figure 4.2: Nuke Showcase, from input images, a point cloud is generated and few lights are added to relight and produce a final image

objects and scenes.

4.1 Relighting a scene

We address the problem of relighting multiple images of a scene taken at the same time of day. No previous solution exists to this problem: Previous methods require multiple lighting conditions [Laffont *et al.*, 2012] or information from a similar scene [Shih *et al.*, 2013].

In the industry Nuke developed by the Foundry integrates a relighting node which allows artists to edit the lighting in a scene. From several input views with a small baseline, a point cloud is generated, and a light added to allow the user to ”relight” the image. No shadow re-computation is involved nor reflectance estimation/manipulation, but the lighting edition is plausible and depends on the quality of the mesh used to approximate the surface (see Fig. 4.2). Despite having a good surface approximation, this single object is relatively easy to reconstruct and Nuke does not allow shadows to be recast with standard CG methods (shadow maps or ray tracing).

In our case, we focus on outdoor scenes composed of multiple objects with complex geometries which cannot be reconstructed properly even by the most recent algorithms. The proxy does not offer good enough quality to use standard cast shadow methods (see Fig. 4.3). Note how the entire roof is approximated by a complete and closed surface in the Monastery scene. In the Villa scene, palm trees are incomplete or approximated by smooth approximate volumes, both of those misestimations can lead to inconvenient artefacts when re-casting shadows. We use the following image formation model:

$$I = Reflectance \times Shading \tag{4.1}$$

$$I = R \times (v_{\text{sun}} L_{\text{sun}} \cos(\omega_{\text{sun}}) + S_{\text{sky}} + S_{\text{ind}}) \text{with, } R \text{ the reflectance,}$$
$$v_{\text{sun}} \text{the visibility,} L_{\text{sun}} \text{the sun color, } S_{\text{sky}} \text{ the sky irradiance and} S_{\text{ind}} \text{the indirect irradiance.} \tag{4.2}$$

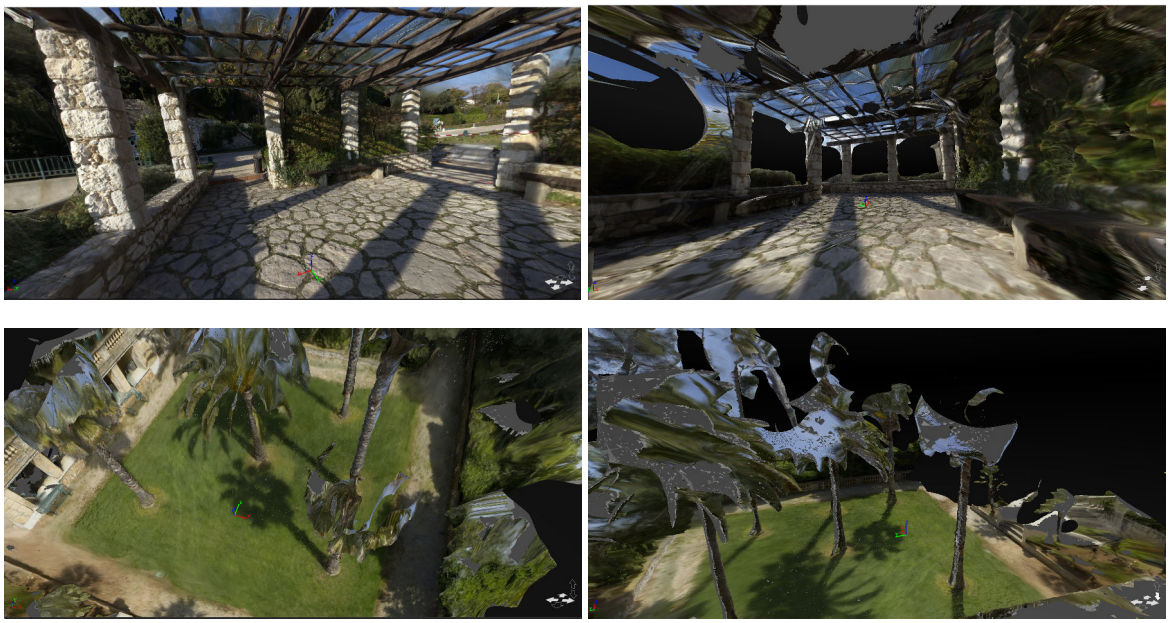


Figure 4.3: Reconstruction artefacts. First row, the monastery scene. Second row the garden of Villa scene.

Figure 4.4: Left, our recast shadow algorithm at initial sun position. Right, result with the reconstructed proxy

Our image formation model derives from the intrinsic decomposition equation 4.1, extended to outdoor scenes as described in chapter 3. This model explicitly separates environment lighting from the sky and indirect irradiance and direct lighting due to the sun (See Eq 4.2). For relighting, while we do not re-evaluate indirect bounce of light we can still manipulate the global ambience of the scene by editing the sky layer (Fig. 4.13). But the main work of this chapter focuses on re-casting shadows for motion of the sun. If the sun direction L_{sun} is updated, the visibility term v_{sun} must be updated for all pixels.

4.1.1 Initial test

Prior to reconstructing a plausible geometry of the shadow caster, a first algorithm was designed to work in image-based space and to try to compensate for the inaccuracy of the geometry. Describing this experience can help preventing similar mistakes in the future but it also inspires the presented algorithm in sub section 4.1.2.

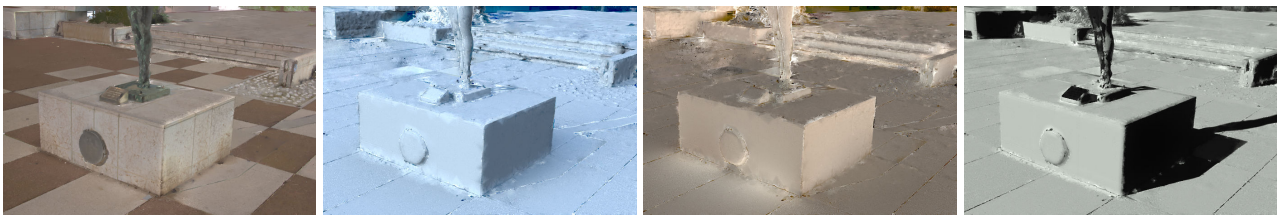


Figure 4.5: Intrinsic decomposition using the method of [Laffont *et al.*, 2013].

This study used the method of [Laffont *et al.*, 2013] to provide the intrinsic decomposition. In this context region in shadow are not identified explicitly so the first step is to introduce a binary classifier. We start by building the histogram of the visibility layer, and identify two peaks corresponding to the shadow cluster and the light cluster. By fitting a normal distribution for each peak, and performing a gradient descent over the histogram to attain a 2 σ limit, approximating the standard deviation, it is possible to obtain a rough estimation of region with a reasonable confidence to be in light or in shadow. The gradient descent operates from the min to the max for the shadow (low intensity) peak and the opposite direction for the light cluster. We thus obtain two clusters defined by the mean and standard deviation of each. This clustering step identifies three regions: black which is in shadow with high probability, white which is in light with high probability and red which is undefined but may be penumbra, self shadowing etc., shown in Fig. 4.6, (left). The clustering is refined and disambiguated by using a Markov Random Field classification solved with an iterated conditional modes **ICM** which takes the two descriptors previously estimated with our clusters as input.

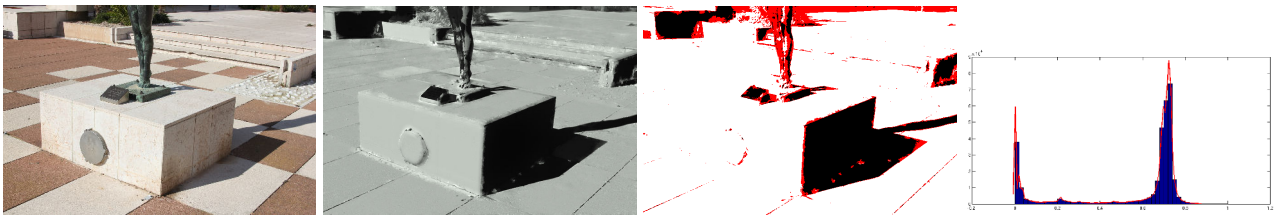


Figure 4.6: Histogram clustering to initialize the MRF.

We focus on manipulating only cast shadow and discard selfshadowing regions by checking the normals which have been propagated using [Levin *et al.*, 2008a] to every pixel and the sun direction

using a mask see Fig. 4.7. For each pixel belonging to a shadow region, casting a ray using the estimated sun direction to intersect the proxy will lead to two cases: an intersection and no intersection. Many pixels do not intersect the proxy due to the poorly reconstructed geometry, particularly near shadow boundaries (Fig. 4.10.

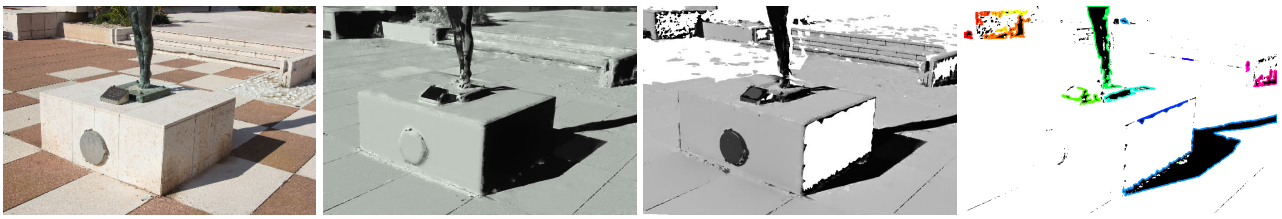


Figure 4.7: Shadow segmentation using the visibility layer of the method of [Laffont *et al.*, 2013].

However intersecting the proxy does not mean the intersection was successful. Indeed, multi view reconstruction algorithms tend to overestimate the geometry, so the distance between the point in shadow and the caster may be inaccurate as well. By storing per pixel the distance between the intersected proxy and the 3D projected points in shadow it is possible to fill pixels in shadow which did not intersect the reconstructed proxy. The filling of these pixels is achieved by minimizing the distance to the caster at pixel p and the weighted average of the distance to the caster at neighbouring pixels (where available). The energy we minimize is:

$$E = \sum_{p}(x_p - \sum_{k=0}^{n} w_k u_k)^2 \tag{4.3}$$

with weight $w = me^{-\nabla d}$, where m is the binary shadow mask so that we only consider regions in shadow, and ∇d is the gradient of the distance to the shadow caster, to preserve continuity of the estimated projected shape of the caster. It is solved by using the standard Matlab backslash operator.

Warping a grid mesh built from the shadow labelling in 2D in a GLSL shader is actually easy to achieve by using the distance to caster as a velocity to control the amount of displacement. However very complex geometries (See Fig. 4.9) such as tree can lead to strong artifacts using this approach since the approximated surface is not curved; the branches introduce strong distance to caster discontinuities that will lead to strange effects during the warp, and the shadow will seems to be compressed.

Many issues at this stage were identified. First the accuracy of the shadow classification and the quality of the intrinsic decomposition are critical. Both of them motivated the need of a new intrinsic decomposition algorithm for outdoor scenes. Second, since the distance to the caster is reconstructed in image-space on a surface which is not reconstructed accurately, a depth to the plane value needs to be assigned. Doing this results in issues along depth discontinuities when composing the shadow along inaccurate object boundaries but also on the ground itself since multi view reconstruction algorithms

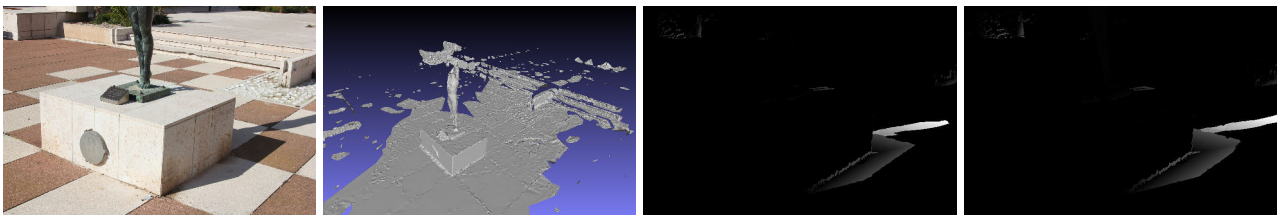


Figure 4.8: Distance to caster propagation for a simple geometry such as a plane or curved surface.

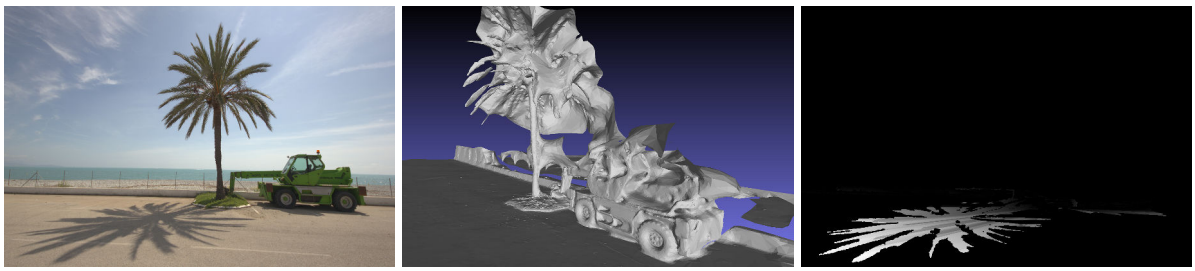


Figure 4.9: Distance to caster propagation for a complex geometry such as a tree.

introduce bumps in the reconstructed mesh. The main reason to create a shadow caster comes from the point that an image based approach is not suitable if all these issues cannot be solved.

4.1.2 Creating a shadow receiver and caster geometry

Recall that shadows cast from the proxy are not accurate enough for relighting, since they do not correspond well to shadow boundaries in the image (see Fig. 4.10). We will move cast shadows approximately by creating a geometric representation of a caster from the shadow boundaries in the original image. The challenge is to disambiguate errors in the proxy so that the shadow produced is as accurate as possible.

While creating caster geometry is related to shape-from-shadow techniques [Savarese *et al.*, 2007], such methods require shadows from multiple light sources. In our case, we only have shadows from a single position of the sun. We thus design an algorithm that preserves the original shadow boundaries in the input image as much as possible and allows some motion of the sun. This process is possible only because we are aware of the shadow region through the shadow classifier presented in the previous chapter 3.

We first enhance the reconstruction of the receiver geometry by assigning to each pixel in the image the depth value of the closest projected 3D point. We found that the resulting depth map, while approximate, results in plausible shadows that we can composite over the reflectance image. We then estimate the geometry of the caster such that it produces shadows that match the shadow boundaries in the original images. We identify the shadow boundaries from the shadow classification layer as

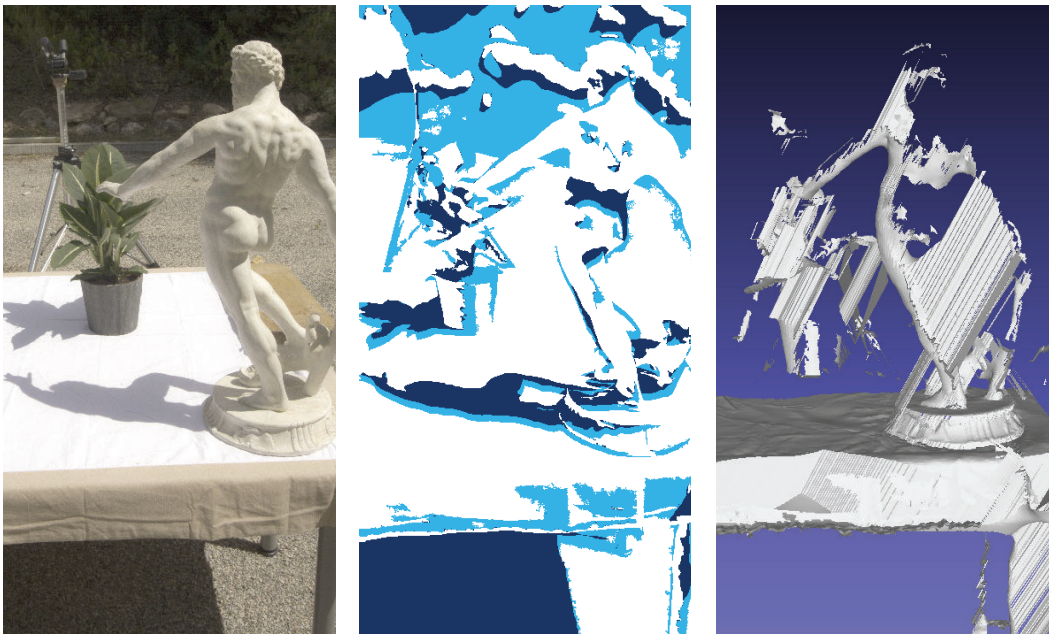


Figure 4.10: Left: Input image. Middle: detected shadow pixels in blue, shadow from the proxy in dark blue. Right: The caster mesh generated from these shadow pixels.

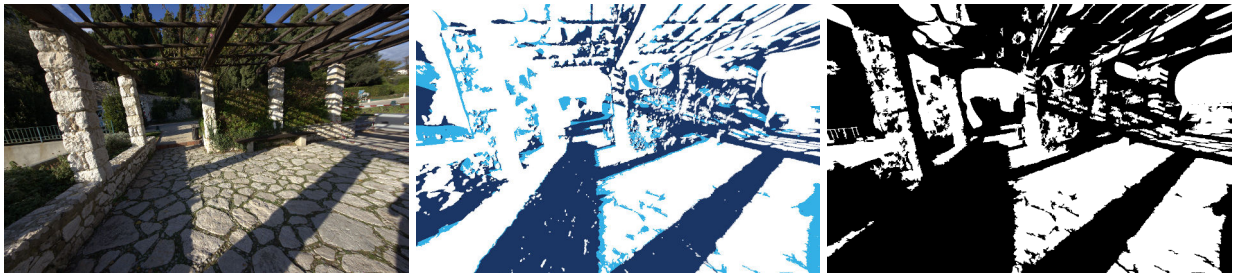


Figure 4.11: Left: Input image. Middle: detected shadow pixels in blue, shadow from the proxy in dark blue. Right : hole filling and small cluster removal performed on the layer. Notice also the shadow of the photographer on the right side which doesn't exist in the proxy.

well as from the propagated v_{sun} layer that sometimes captures fine details lost by the binary classifier (Fig. 4.11). We consider pixels to be in shadow if pixel p is classified as shadow, or if $v_{\text{sun}}(p) < \tau_s$. We used $\tau_s = 0.8$ for all our results. To estimate a 3D caster position at each shadow pixel, we shoot rays in the direction of the sun θ_{sun} and record the distance of the closest intersection with the 3D proxy. Pixels for which the ray does not intersect the proxy receive the distance of the nearest valid pixel.

We triangulate the shadow pixels in image space to create a mesh that we lift in the direction of the sun using the recorded distance from the cast point. Fig. 4.10 illustrates the resulting 2.5D caster which re-creates the shadow boundary in the image.

Incorrect reconstruction and numerical imprecision can result in erroneous triangles in the caster that partly re-project on lit pixels. We remove such triangles by visiting all pixels in light and casting

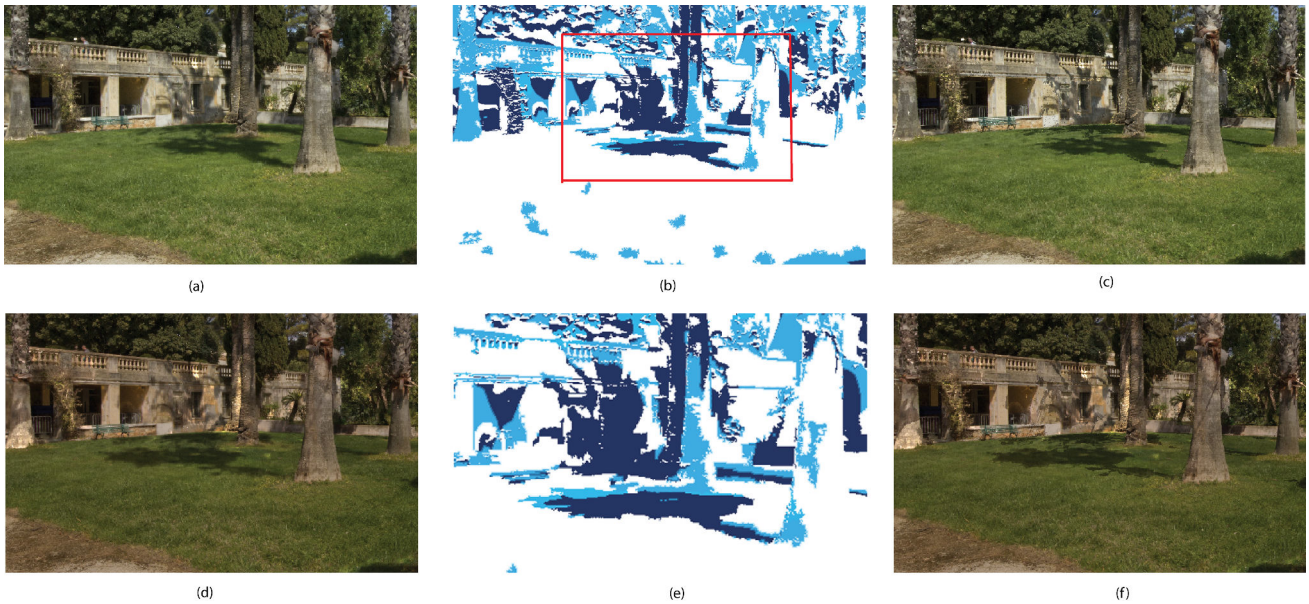


Figure 4.12: Our result at the original sun position (a) and our relighting result for sun motion (d). Our result without performing morphological operation and gaussian blur to clean our layer in (c) for original sun position and (f) for some motion of the sun. (b) and (e) represent our shadow layer with detected shadow pixels in blue, shadow from the proxy in dark blue.

rays in the sun direction. If a triangle of the caster mesh is intersected by more than ϵ such rays, it is removed. We used $\epsilon = 3$ for all our results. Our shadow labeling also sometimes mis-classifies pixels as shadow in small regions as shown in Fig. 4.12. To filter these errors we cluster the pixels in shadow and remove small clusters with less than 100 pixels and clusters for which less than 30% of the pixels yield an intersection with the proxy. We adjust the reflectance of such pixels to bring them in light using the v_{sun} layer. The difference of quality with and without this step can be observed in Fig. 4.12.

4.1.3 Moving shadows and adjusting Shading

To get coherence with the real movement of the sun for a particular location, we use the GPS coordinates and EXIF tags from the captured shots to estimate the sun motion. To move shadows, we simply update the sun direction θ_{sun} and trace rays from each pixel in that direction. We compute intersections against the caster using the ray tracing Intel *embree* library, which provides interactive feedback for the images (see Fig.4.13).

However, our caster geometry only reproduces the shadows captured in the image. As a result, discontinuities can appear when the shadow is moved away from the border. To reproduce a plausible visibility layer, we complete the missing shadow in these areas using the shadow of the proxy geometry, then use morphological filters with a sequence of closing and opening operations [Soille, 2003], to remove small holes. Finally we apply a small Gaussian blur on this new shadow layer to mimic soft

shadows and to fill small holes caused by disconnected triangles in the caster mesh. Of course, this approach is limited in terms of sun motion and quality; we describe such issues in the following result section. Note that a shadow caster could be drawn by an artist to get an high quality layer for a larger motion of the sun which still requires a high quality intrinsic decomposition, but our focus here is to demonstrate what can be done with an automatic method.

Having plausible shadows is good fist step but we still require good quality shading without accurate normals. To do so, we use the most reliable information available, i.e., the intrinsic layers, rather than the inaccurate proxy. By scaling them in the most plausible way, we can render plausible images and still be consistent with sun motion. To do so, we update all layers in a different way. We approximate the effect of sun illumination changes by adjusting the sun shading intensity w_{Sun} according to a cosine factor with respect to elevation and the horizontal plane, and shifting S_{sky} towards red in the morning and afternoon using a weight w_{sky}. To maintain the illusion of shading change without recomputing indirect bounces, w_{ind} is diminished by a similar amount. Finally, we detect sky pixels in the input image and change their color near the horizon and recompute them using the input image itself.

The sky detector works in real time in a GLSL shader. We select all pixels above the horizon from the input image using color information which have a higher color ratio for the blue channel ($blue \geq red + \epsilon, blue \geq green + \epsilon$ with $\epsilon = 0.05$). The algorithm is described in Alg.1.

We define the following quantities: t_{Noon} to be noon 12:00, ∇_{Noon} the difference of time between the captured images and noon, t_{sunset} sunset time, ∇_{sunset} the difference of time between the captured image and sunset.

Algorithm 1 Shading code

if $t_{Noon} \geq \nabla_{Noon}$ **then** ▷ Make the sun brighter as we are getting closer to noon
  $\alpha SunIntensity \leftarrow 1 - (\frac{t_{Noon}}{\nabla_{Noon}})$
  $w_{Sun} \leftarrow 1.0 + \alpha SunIntensity$
else ▷ Make the sun darker
  $\alpha_{SunIntensity} \leftarrow 1 - \min(1, \frac{(t_{Noon} - \nabla_{Noon})}{(t_{sunset} + \nabla_{sunset})})$
  $w_{Sun} \leftarrow 1.0 - \alpha SunIntensity$
  if $t_{Noon} > \nabla_{Noon} + t_{sunset}$ **then** ▷ Make the sky red
    $\alpha sky \leftarrow \frac{t_{Noon} - \nabla_{Noon} - t_{sunset}}{\nabla_{sunset}}$;
    $w_{sky} \leftarrow vec3(1, \alpha_{sky}, 0.8 * \alpha_{sky})$;
    $w_{ind} \leftarrow *w_{sky} * scale_{indirect}$;
  end if
end if

▷ Apply weights to our shading model

$out_{color} \leftarrow reflectance * (w_{ind} * indirect + w_{sky} * sky + (corr_{Vis} * w_{Sun} * shadow) * sun)$;
$out_{colorskypixel} \leftarrow Input_{img} * (w_{sky} * horizon_y)$;

Figure 4.13: Input image relight in our interactive system in Villa Eilen Roc.

4.2 Results

4.2.1 Single Image

We show results for relighting for the Villa scene in Fig 4.13, for Street in Fig 4.14, for Monastery in Fig 4.15. See Section 5.1 for information on these scenes. Fig 4.13 shows the same view relit at different time of the day.

4.2.2 Ground truth

We captured several lighting conditions for the Statue and Plant scene, to allow a ground truth comparison. We only used multi-view capture of the central image (i.e., a single lighting condition) for all intrinsic decomposition and relighting computations. We show and discuss the results in Sec 5.7.

4.2.3 Image Based Rendering

Our relighting approach can also be used for image-based rendering and changing lighting conditions. In Fig. 4.18, we show some samples of view interpolation and free-viewpoint navigation path in the Villa dataset in which we use the algorithm of [Chaurasia *et al.*, 2013]. We record the path, change

Figure 4.14: Input image relit in our interactive system for the dataset taken in Saint Paul de Vence. Top with reconstructed proxy, bottom our result.

lighting conditions and play it back with the new illumination, since all the input images used for IBR have been updated, no modification of the original IBR algorithm is required. This is best appreciated in the accompanying video.

4.2.4 Compositing

We show a virtual object composited using our layers in Fig 4.19.

4.3 Conclusion and future work

We presented our automatic method introducing multi-view relighting, and demonstrate its utility for image based techniques with illumination changes from the single image case to image based rendering approaches. Our method is the first to allow relighting by manipulation of shading and cast shadows with good quality. Our approach opens up large possibilities for future work. Currently, our relighting system is limited to outdoor scenes with sunlight and well-defined cast shadows which can be segmented. For outdoor scenes with overcast sky, the problem can appear simpler but it is different. Since the variation between shadow and light is much smoother, any errors in the initial lighting condition

Figure 4.15: Input image relit in our interactive system for the dataset taken in a garden of Nice, near the Cimiez monastery. Top with reconstructed proxy, bottom our result.

estimation will not get corrected by our re-estimation using the shadow classifier; a new algorithm needs to be proposed as mentioned in the previous chapter 3. Manipulating hard shadows requires segmenting them to reconstruct a shadow caster to replace the proxy. Currently we can achieve this only for small motion of the sun. Another direction for future work is also related to the intrinsic decomposition model, the development of a more complete image formation model which can incorporate non-diffuse behaviour and allow manipulation of reflections for objects in the initial scene and also for inserted virtual objects.

Figure 4.16: Input images used for the image based rendering path.

Figure 4.17: Input images relit used for image based rendering path.

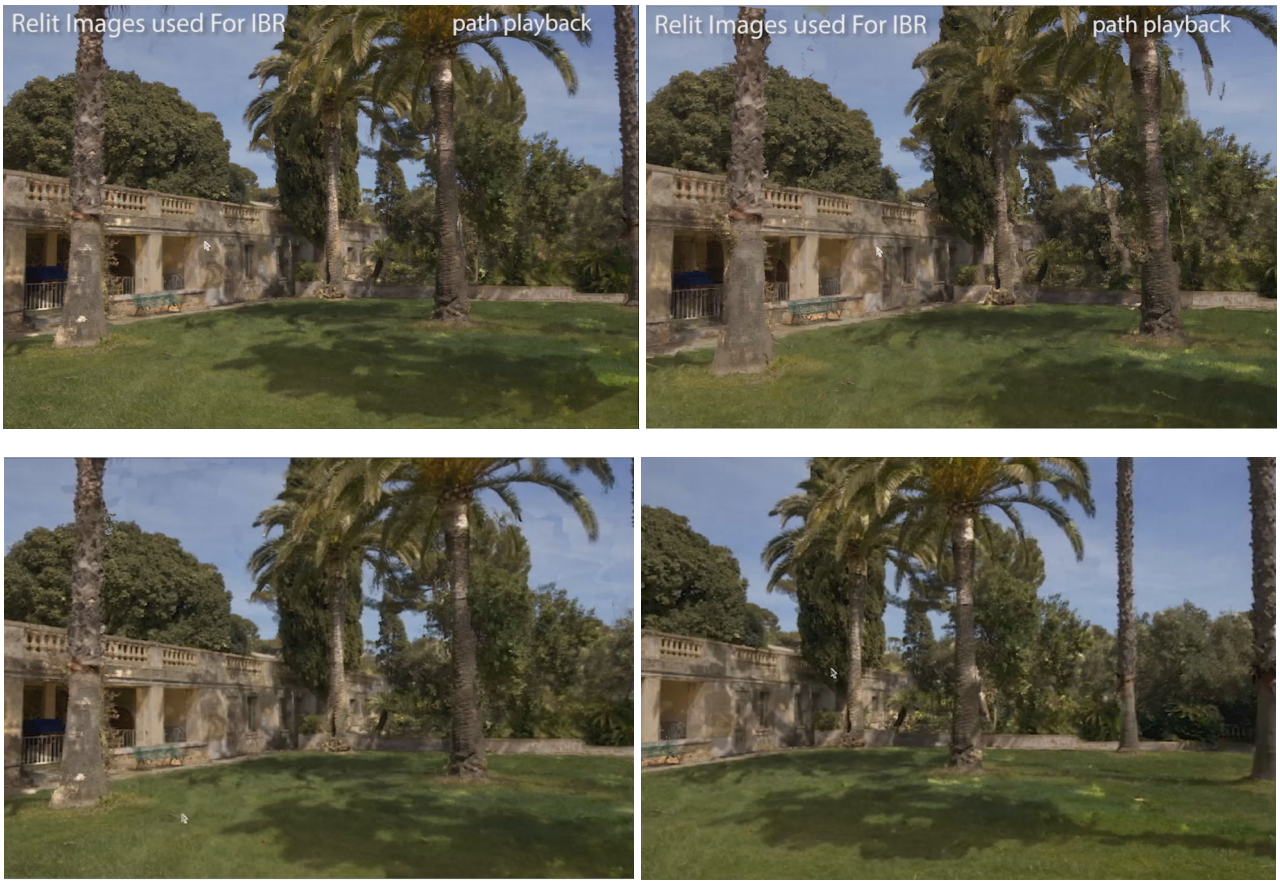


Figure 4.18: Interpolated images using the depth superpixel warp approach [Chaurasia *et al.*, 2013].

Figure 4.19: An example of CG insertion using the villa scene dataset and its environment lighting estimation to insert a creature.

Chapter 5

Evaluation

The purpose of this chapter is to evaluate several steps of the algorithms introduced in chapter 3 and chapter 4. It contains :

- A description of the real world scene used to evaluate the method. It also contains the environment map fitted, and a selection of the captured images is provided in the intrinsic decomposition result subsection.
- A comparison of our automatic sunlight calibration and environment map estimation with the method of [Laffont *et al.*, 2013], which uses a grey card and chrome ball Sec. 5.2.
- A visual comparison of our algorithm with state-of-the art intrinsic image methods and shadow classifiers Sec. 5.3. This part contains the result of a selection of input images for each dataset.
- An evaluation of the robustness of our approach by decreasing the number of input images used, Sec. 5.5.
- A ground-truth quantitative evaluation of our algorithm and comparison to [Laffont *et al.*, 2013], Sec. 5.6.
- A ground-truth comparison of our synthetic relighting with real photographs taken at different times.

Table 5.1 details the number of images used for each scene, along with the number of vertices of the proxy geometry.

	Street	Monastery	Villa	Statue	Toys
#images	61	61	138	60	73
#proxy	2Mi	2.2Mi	6Mi	4.6Mi	4.6Mi

Table 5.1: Number of input images and number of vertices of the estimated 3D proxy, for each dataset.

5.1 Scenes description

5.1.1 Street

Figure 5.1: Google Earth view of the street scene captured in Saint Paul de Vence. Latitude $43°41'50.68"N$ and longitude $7°7'17.09"E$. Our synthesized environment map in top right. Bottom left : our mesh with color transfer using the median. Bottom right : normal of the reconstructed mesh.

Figure 5.2: Input image (a), shadow from inaccurate 3D model (b), input point cloud (c), cast shadow in our user interface to setup the sun direction, blue and green represent region in shadow (d).

This scene highlights the difficulty to classify as shadow or in light, region of cast shadows by poorly reconstructed geometry like the cactus. The inaccuracy of the reconstructed model directly affects the estimation of the irradiance collected per point in this region which has a strong impact for the two steps of our method, the sun calibration and then the shadow classification.

5.1.2 Monastery

Figure 5.3: Left: Google Earth view of the Monastery scene captured in the Cimiez area of Nice. Latitude $43°43'7.40"N$ and longitude $7°16'43.57"E$. Right: our synthesized environment map.

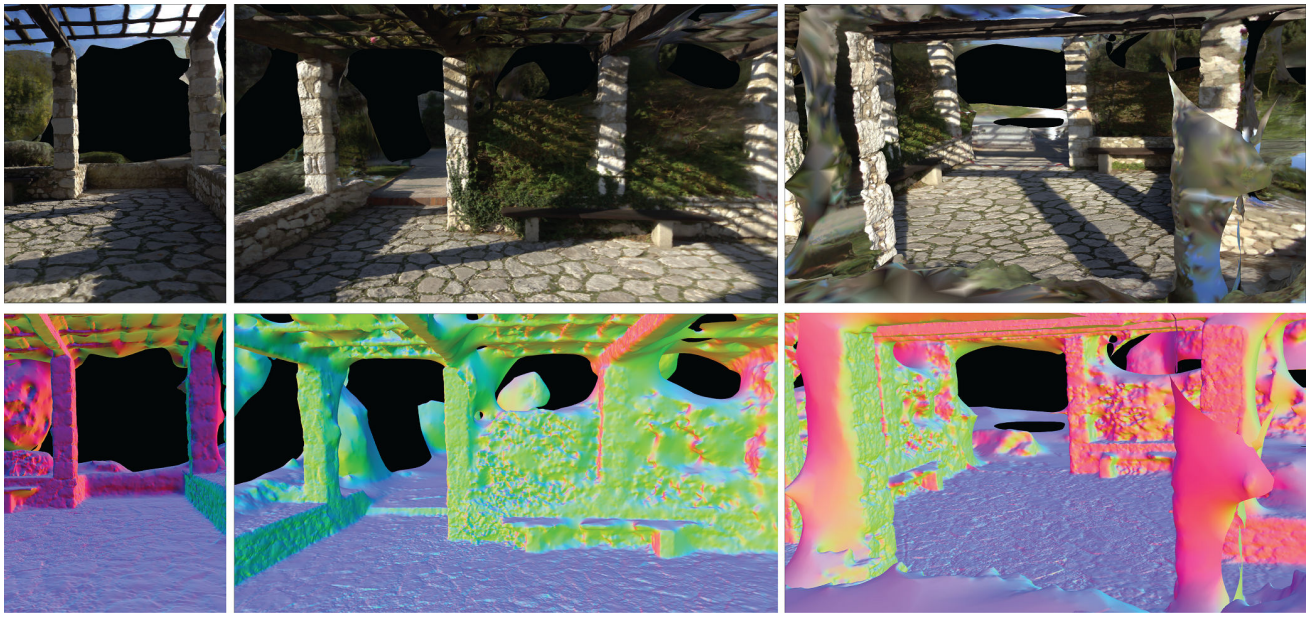


Figure 5.4

The shadow classification of this scene is very challenging since the ground is made of stones and jointing mortar which can perturb the image based pairwise cost function described in chapter 3. The dataset is also included in the shadow labelling comparison in section 5.11. The inaccuracy of the reconstructed 3D model also motivates the need of an explicit shadow classification to relight the scene especially because of the grid-like ceiling which cast a complex shadow in the captured image but cast a massive shadow in the case of a relighting method which uses the reconstructed proxy since the ceiling is reconstructed as if it was filled. However, we also discuss the limit of our method to very fine structure elements which are really hard to segment before being classified.

5.1.3 Villa

Figure 5.5: Left: Google Earth view of the garden of the Eilen Roc Villa scene in near Antibes. Latitude $43°32'42.25"N$ and longitude $7°7'49.98"E$. Right: our synthesized environment map.

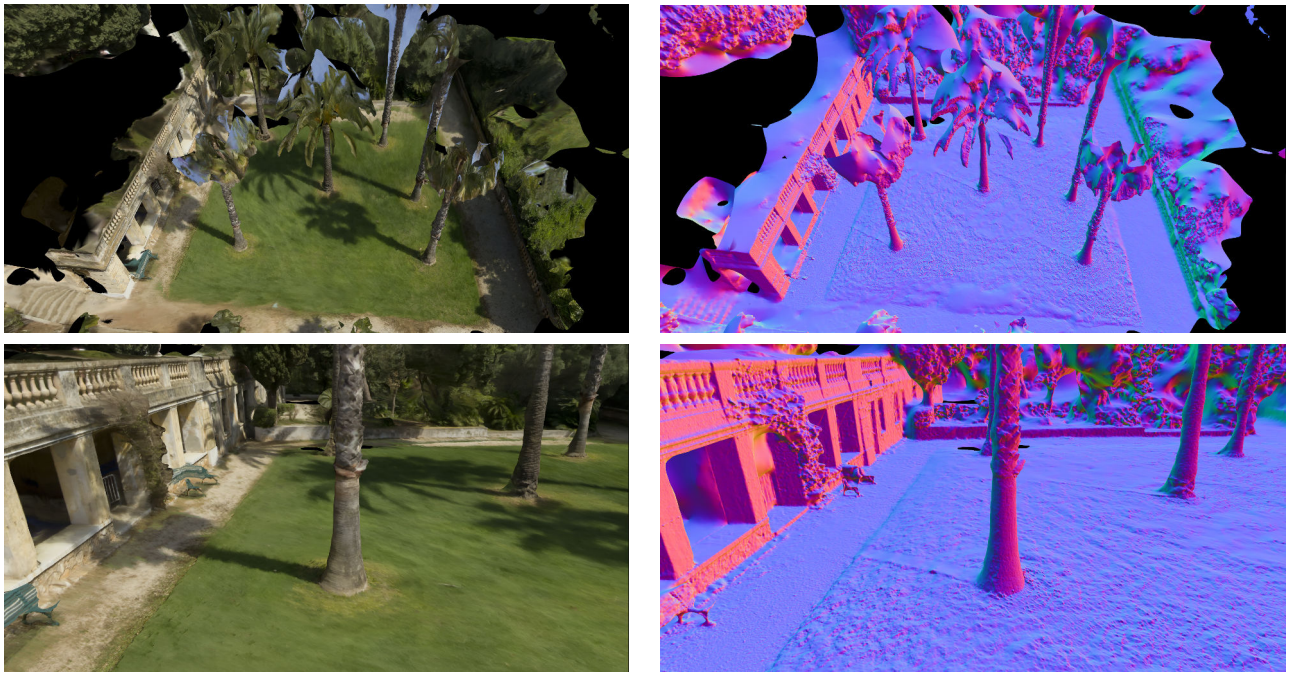


Figure 5.6: Color proxy and normal rendering of the reconstructed geometry.

This dataset was originally captured with 138 images with a specific camera path suitable for use with on image based rendering method described in [Chaurasia *et al.*, 2013]. The trunk of trees and the ground are reasonably well reconstructed, however we can observe that the palms trees are not. This scene was also used to observe the side effects of geometry accuracy on our pipeline by reducing the number of input images used for reconstruction in the section 5.5.

5.1.4 Plant and statue

Figure 5.7: Google Earth insight of the statue and plant scene captured near Inria Sophia Antipolis. Latitude $43°36'51.54"N$ and longitude $7°4'6.28"E$.

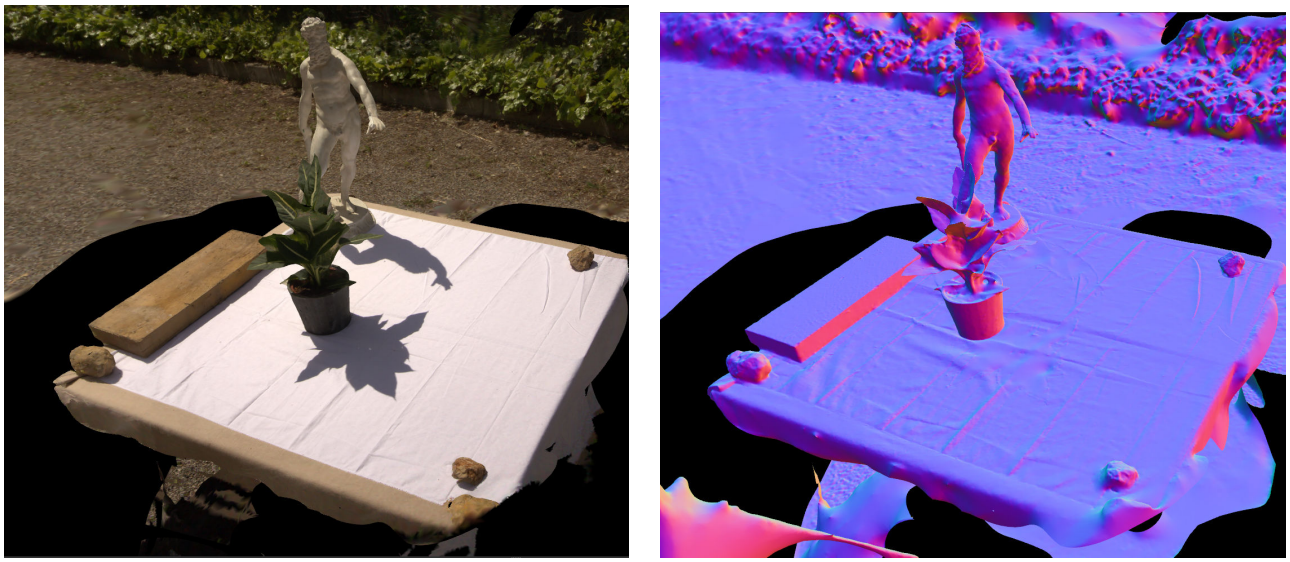


Figure 5.8: Color proxy and normal rendering of the reconstructed geometry.

This dataset is used in two comparisons with previous work for intrinsic decomposition Sec. 5.4, Fig. 5.12 and the shadow classification Sec. 5.3, Fig. 5.11. All previous methods failed for both comparisons in this case, which is particularly surprising at least for the tablecloth. Compare to others scene, the diversity of reflectance in this scene is very low.

5.1.5 Toys

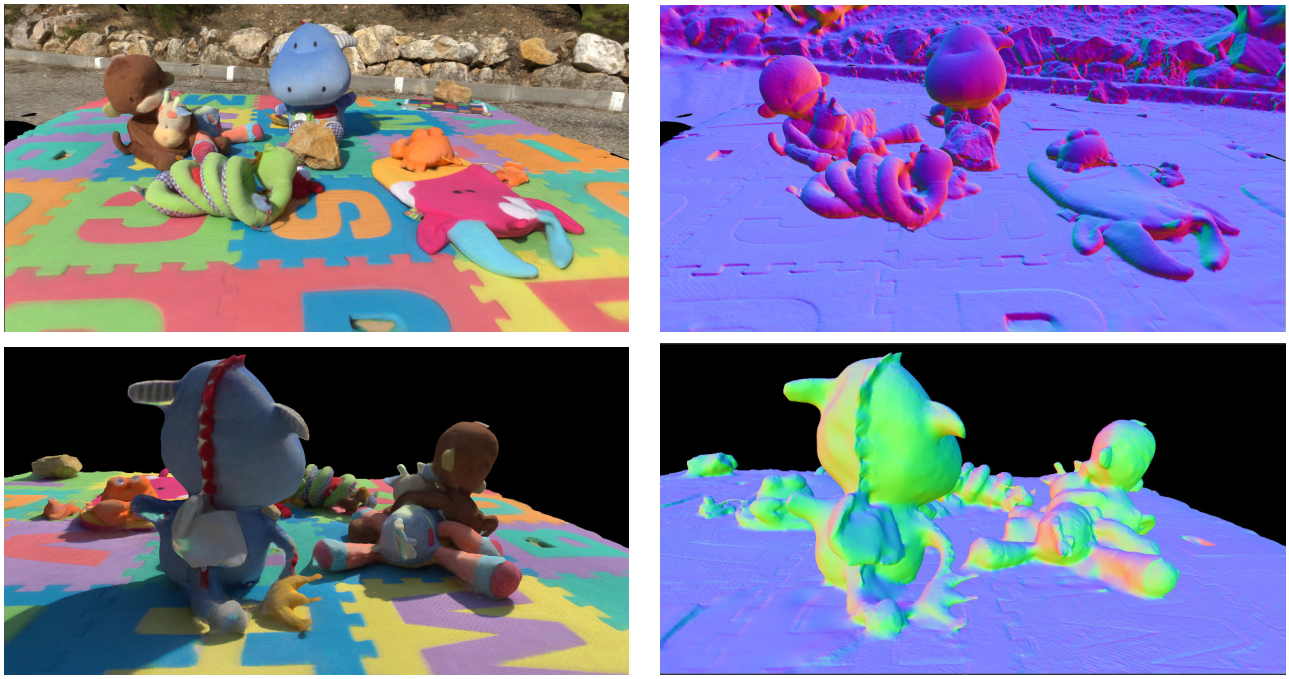


Figure 5.9: Color proxy and normal rendering of the reconstructed geometry.

For this final dataset ground truth and our synthesized environment map are provided as well in the sun calibration section Sec. 5.2. The clipping range of interest in this scene is small, like the statue and plant case Sec. 5.1.4 but in this case a wide diversity of reflectances is observed. For these reasons, the dataset is involved in three comparisons with previous work for intrinsic decomposition Sec. 5.4, the shadow classification Sec. 5.3, Fig. 5.11 and the sun calibration Sec. 5.2.

Figure 5.10: Comparison using a chrome ball and grey card (left) and our synthesized environment map with automatic calibration (right). Although our environment map misses details on the ground and horizon, it captures the overall color distribution of the ground and sky, yielding reflectance results (lower row) visually similar to the ones obtained with additional information.

5.2 Sun calibration

Recall that, compared to [Laffont *et al.*, 2013], all steps in our approach are automatic, removing the need for the chrome ball, grey card, parameter setting and inpainting steps. Figure 5.10 provides a comparison between our automatic decomposition and a downgraded version of our algorithm where we used the captured chrome ball and grey card calibration of [Laffont *et al.*, 2013]. Our calibration estimates a sun color of $(2.7, 2.3, 2.4)$ while the grey card yields $(2.7, 2.7, 2.7)$. Our estimated environment map captures the overall color distribution of the sky and ground and results in a reflectance on par with the one obtained with a chrome ball and manual calibration.

5.3 Shadow classifier

Figure 5.11 shows a comparison with two single-image shadow classifier methods [Zhu *et al.*, 2010] and [Guo *et al.*, 2011]. Our classifier works well in most cases, and compares favorably to the previous approaches. The method of [Guo *et al.*, 2011] often gives very good results (last row), but can sometimes reports false positives or fails (top row).

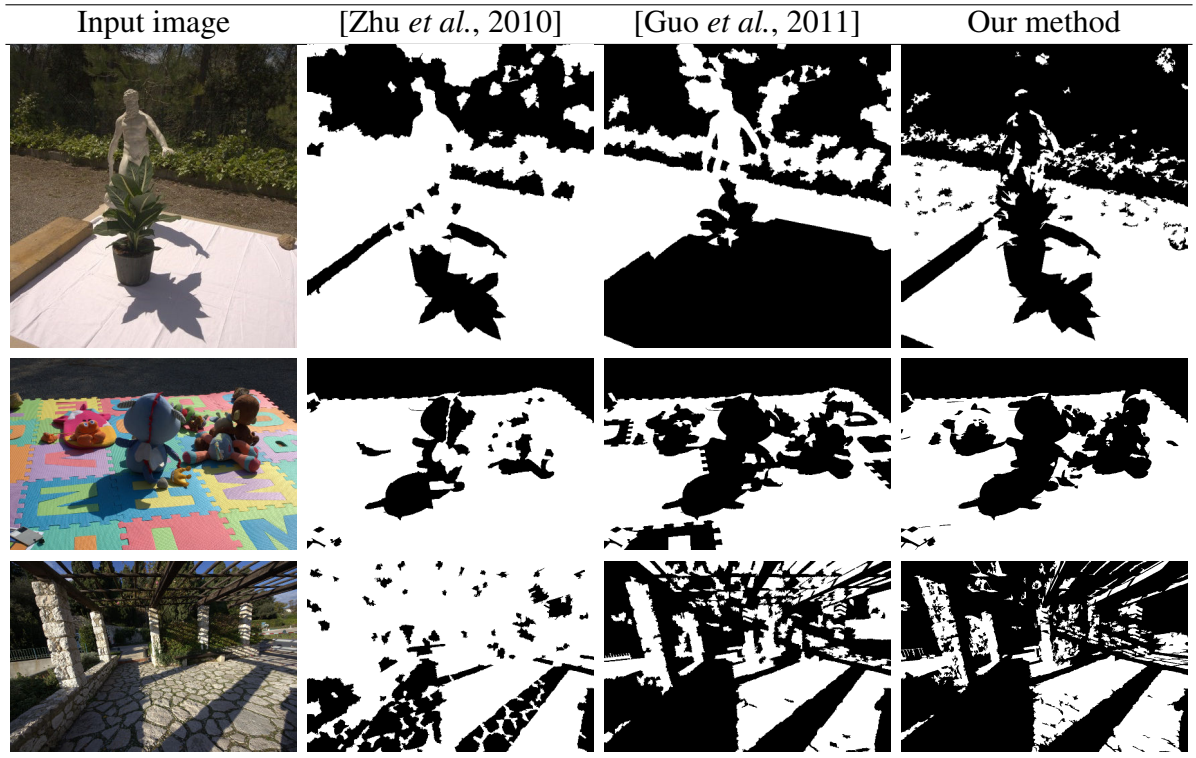


Figure 5.11: Comparison with existing shadow classifiers. [Zhu *et al.*, 2010] misses shadow details while [Guo *et al.*, 2011] tends to produce false positives. Our approach leverages 3D information to avoid such errors.

Note that in section 5.5, we provide the result of the shadow classification of the same scene reconstructed with varying numbers of input views to observe the effect of the geometry accuracy.

5.4 Intrinsic decomposition results

5.4.1 Real world scenes

We provide our results on a selection of real-world scenes presented in Section 5.1 :

- toys scene in Fig. 5.14.
- street scene in Fig. 5.15.
- plant and statue scene in Fig. 5.16. and Fig.5.17.
- villa[1] scene in Fig. 5.18
- monastery scene in Fig. 5.19.

5.4.2 Comparison

We compare with recent state of the art intrinsic image algorithms, namely two single-image approaches [Barron and Malik, 2013b; Chen and Koltun, 2013] which also use depth information. From the results presented in these papers, these methods outperform previous single-image solutions which are typically derived from the Retinex algorithm. We also compare to the multi-view method of [Laffont *et al.*, 2013]. We used the original code of these papers, and reported results to the authors who ensured that we set parameters correctly.

We present two test scenes for comparisons in Fig. 5.12 and Fig. 5.13. The Plant and statue is a simple scene described in Sec.5.1.4, with a cast shadow on a tablecloth. The proxy reconstruction is of quite high quality except for the plant. From the results we can clearly see that the single image methods are not suited to outdoor scenes with cast shadows, and there is always a residue in the reflectance layer. Our algorithm benefits from the better 3D reconstruction provided by multi-view stereo. The method of [Laffont *et al.*, 2013] also has some residue due to their use of approximate non-binary visibility values that tends to compensate for errors in the estimated shading. By enforcing binary visibility we obtain robust shadow classification, and consequently correcting reflectance across shadow boundaries in a reliable manner, our method produces better results overall.

In Fig. 5.13, the single image methods [Chen and Koltun, 2013; Barron and Malik, 2013b] both have residues in the reflectance. The method of [Laffont *et al.*, 2013] has similar results with ours for this scene: ours has slightly less residue in reflectance, but does miss-classify some of the checkerboard colors as shadow. In addition, that method overestimates indirect light in corners with inaccurate reconstruction, which we attenuate with the cosine factor. It is important to recall again that the method of [Laffont *et al.*, 2013] is not fully automatic, requiring several manual steps described in

[1]The movie Magic in the Moonlight of Woody Allen was filmed partially in the Villa Eilenroc, Antibes.

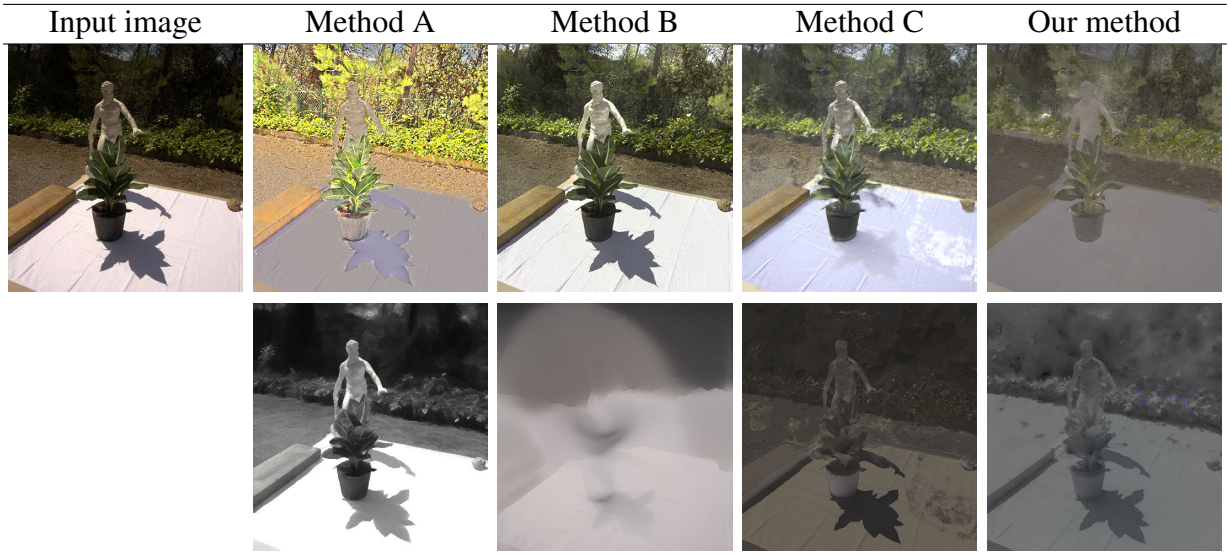


Figure 5.12: Comparisons with existing intrinsic image methods, reflectance and shading respectively top and bottom row. Results are shown with scale factor and gamma-correction. Our approach removes the hard shadow, which allows us to subsequently relight the scene. Method A, [Chen and Koltun, 2013]. Method B, [Barron and Malik, 2013b] . Method C, [Laffont *et al.*, 2013].

the main text. We manually picked the best result of their method to compare with.

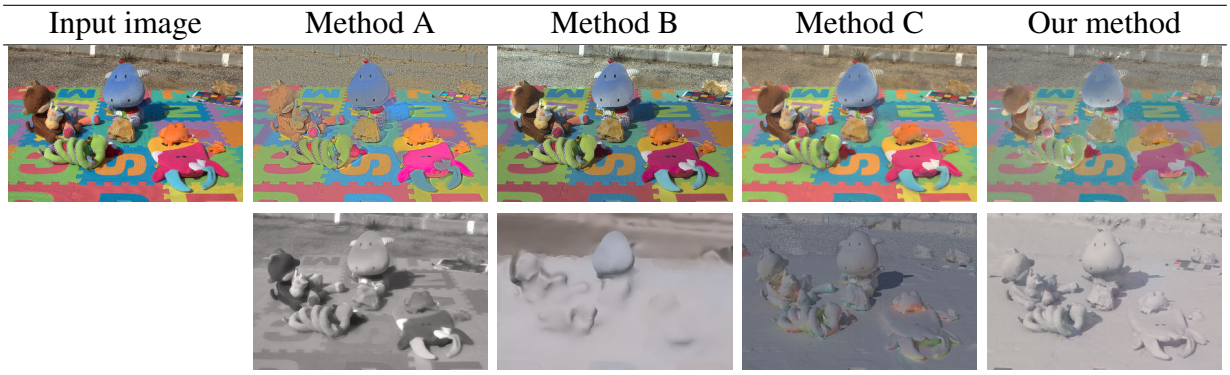


Figure 5.13: Comparisons with existing intrinsic image methods, reflectance and shading respectively top and bottom row. Results are shown with scale factor and gamma-correction. Method A, [Chen and Koltun, 2013]. Method B, [Barron and Malik, 2013b] . Method C, [Laffont *et al.*, 2013].

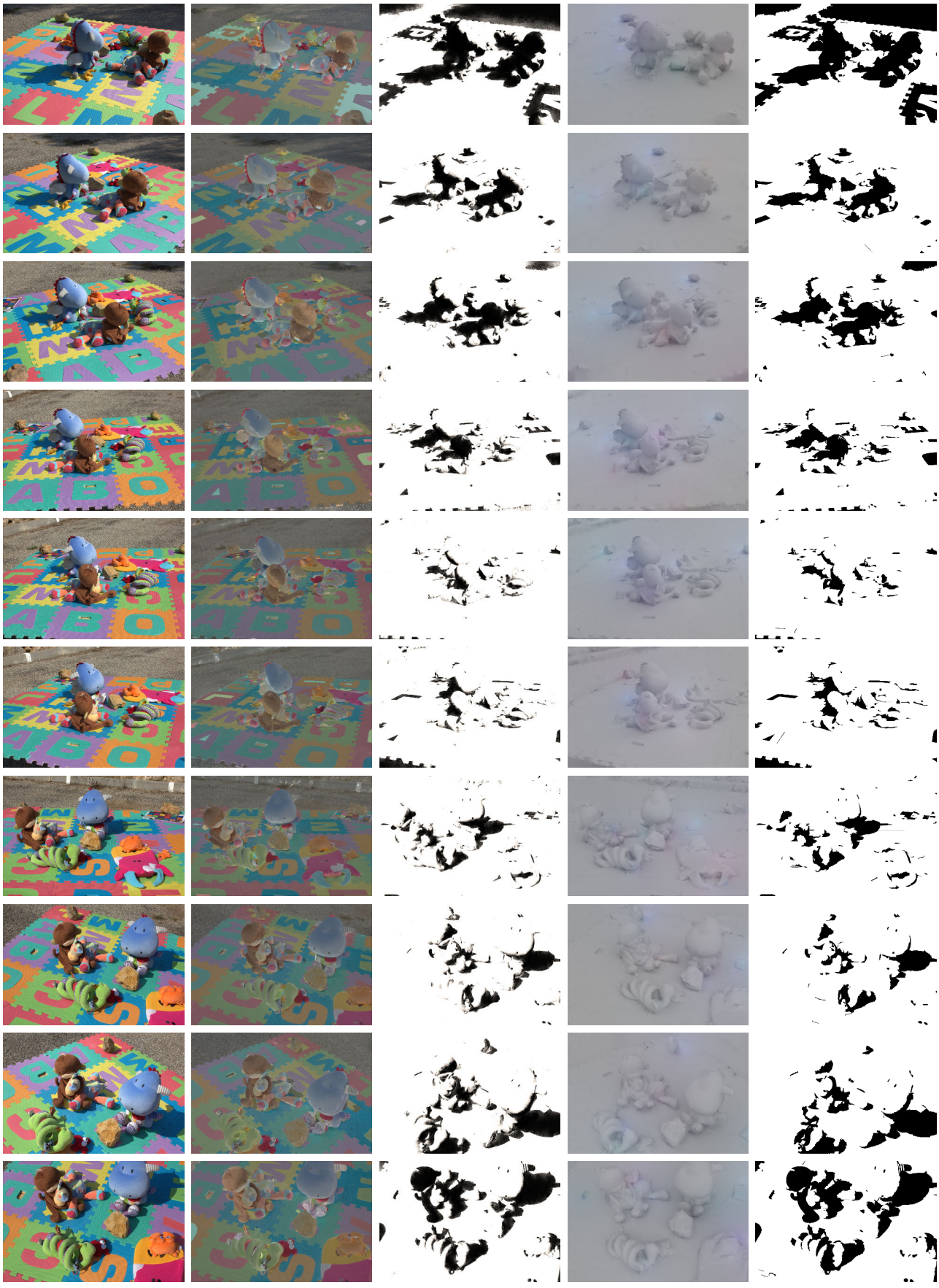


Figure 5.14: Intrinsic decomposition results for the toys scene.

Figure 5.15: Intrinsic decomposition results for the street scene.

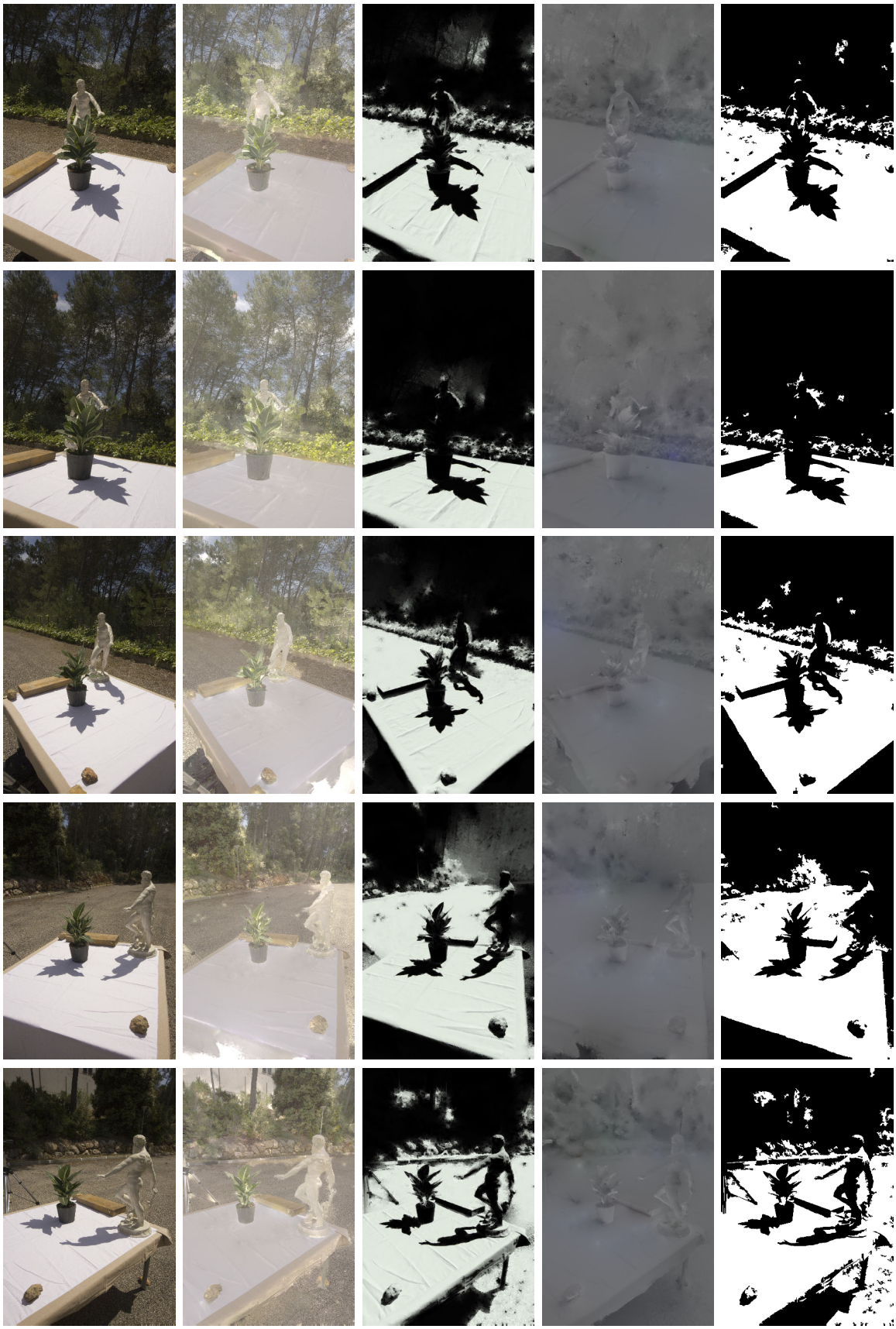


Figure 5.16: Intrinsic decomposition results for the plant scene.

Figure 5.17: Intrinsic decomposition results for the plant scene.

Figure 5.18: Intrinsic decomposition results for the villa scene.

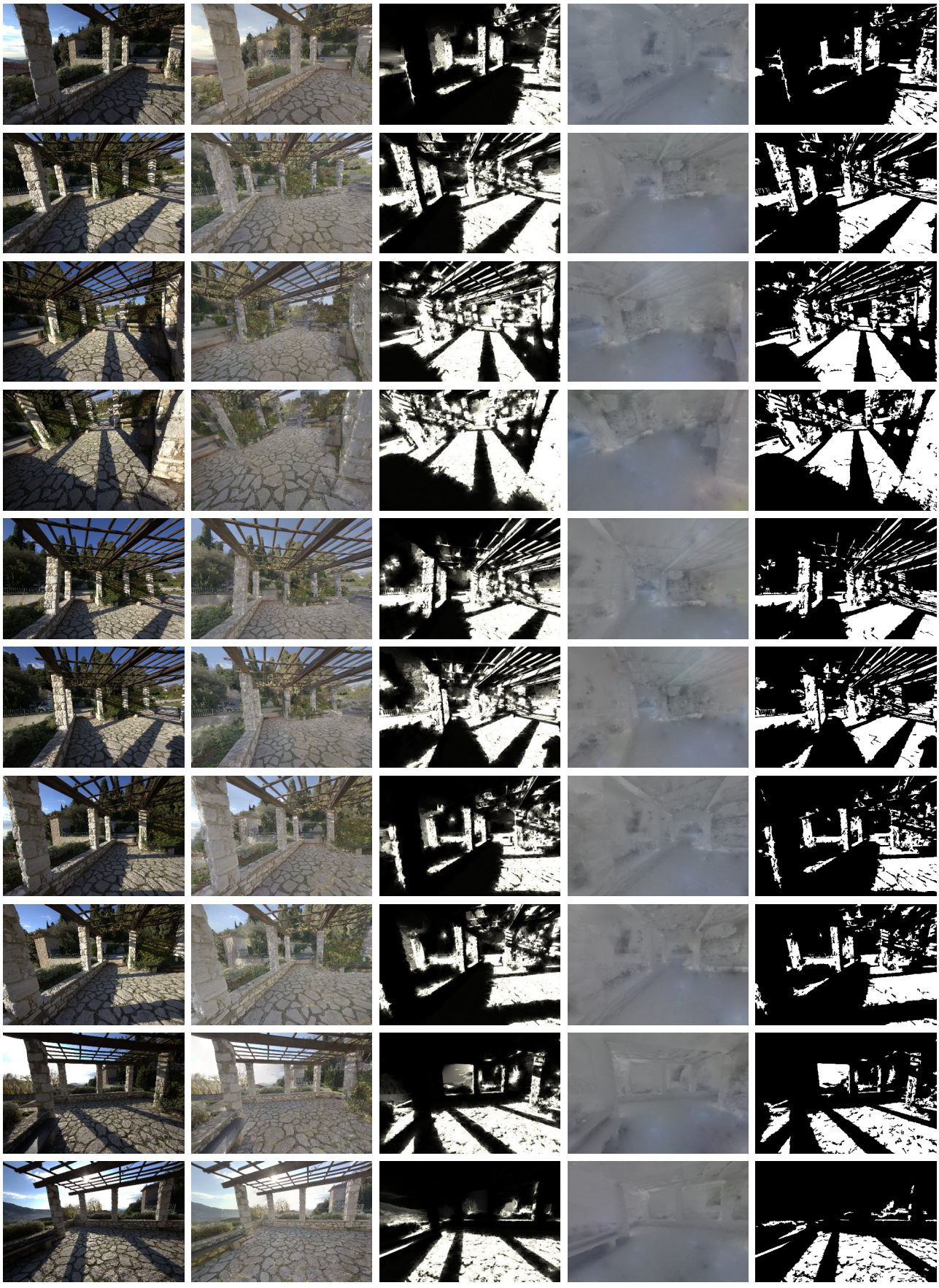


Figure 5.19: Intrinsic decomposition results for the monastery scene.

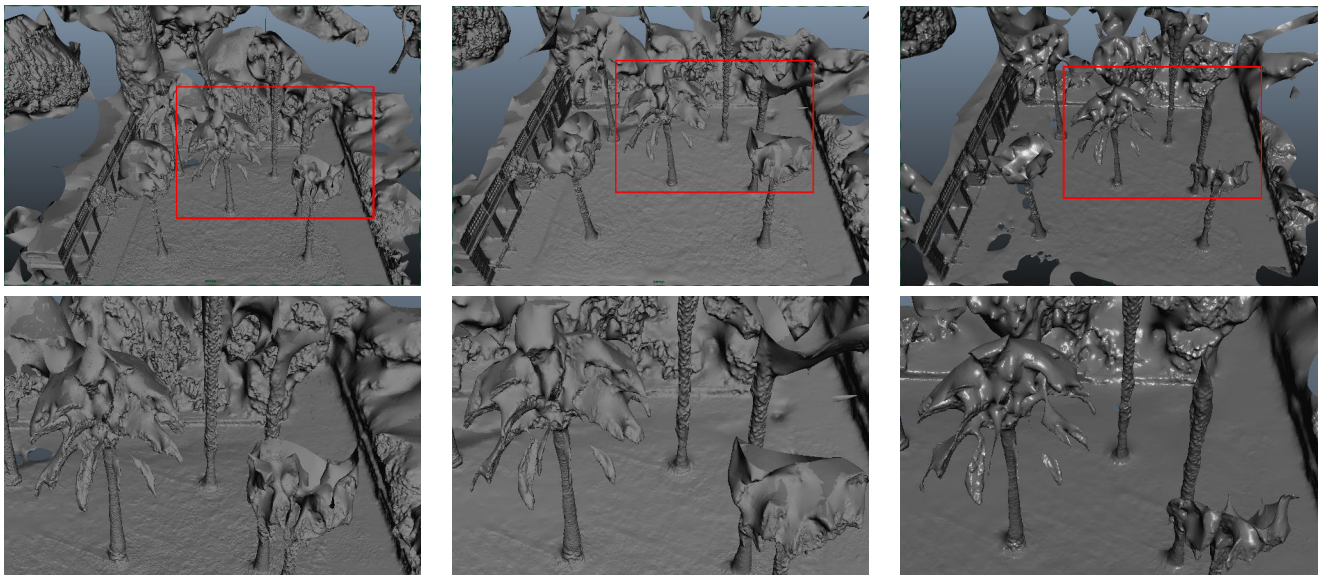


Figure 5.20: 3D reconstruction with 138, 68 and 34 views. The reconstruction is increasingly incomplete as we lower the number of images. See Fig. 5.21 for the corresponding intrinsic decompositions.

5.5 Effects of geometry accuracy

As is often the case with multiview stereo reconstruction, we found it easier to capture a large number of images rather than attempting to find the smallest set of images that would be sufficient to run our method.

In theory, lowering the number of input images can impact several aspects of our pipeline. First, using fewer images results in fewer samples to estimate the diffuse radiance of the proxy geometry Sec 3.2 and fewer candidate pairs for sun calibration Sec 3.3.

We conducted a small experiment to evaluate the practical impact of the number of input images on the quality of the end decomposition. Figure 5.21 shows that despite reducing the number of images from 138 to 34 our algorithm produces consistent results. This success is due to the fact that our shadow labeling algorithm leverages image information to identify accurate shadow regions even when the shadow caster is not well reconstructed, as shown in Figure 5.20.

Figure 5.21: Decreasing the number of input images does not have a significant impact on the quality of the decomposition. See Fig. 5.20 for the corresponding 3D reconstruction. The scene was reconstructed using 138, 68 and 34 views. We show the input image, reflectance, lighting and shadow classifier results for these 3 cases.

Figure 5.22: Decreasing the number of input images does not have a significant impact on the quality of the decomposition. See Fig. 5.20 for the corresponding 3D reconstruction. The scene was reconstructed using 138, 68 and 34 views. We show the input image, reflectance, lighting and shadow classifier results for these 3 cases.

5.6 Ground truth

Figure 5.23: Overview of synthesized images for our ground truth evaluation.

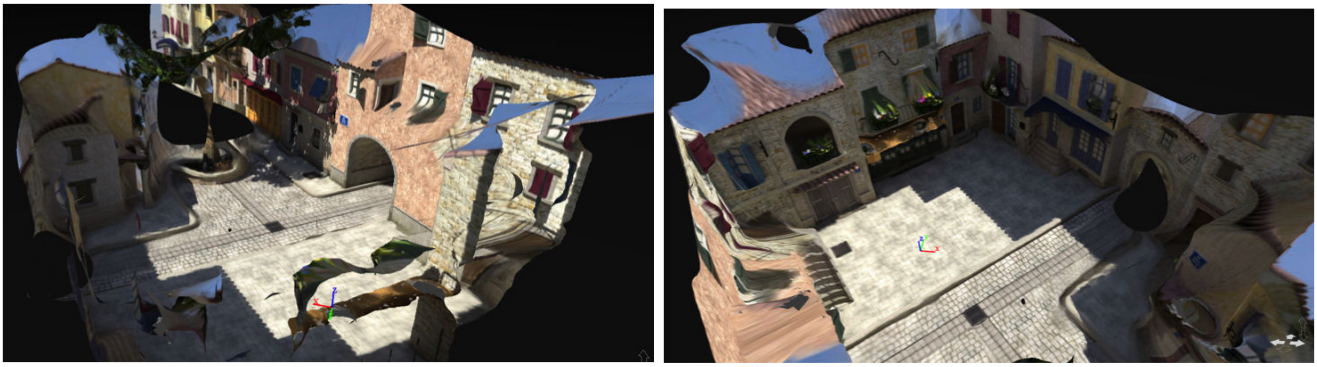
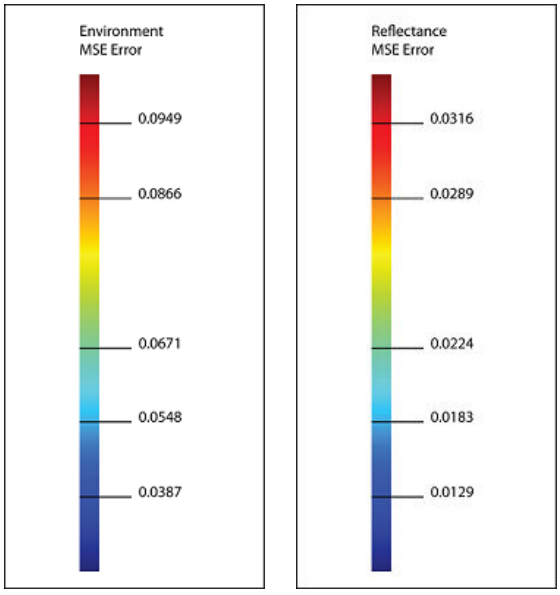


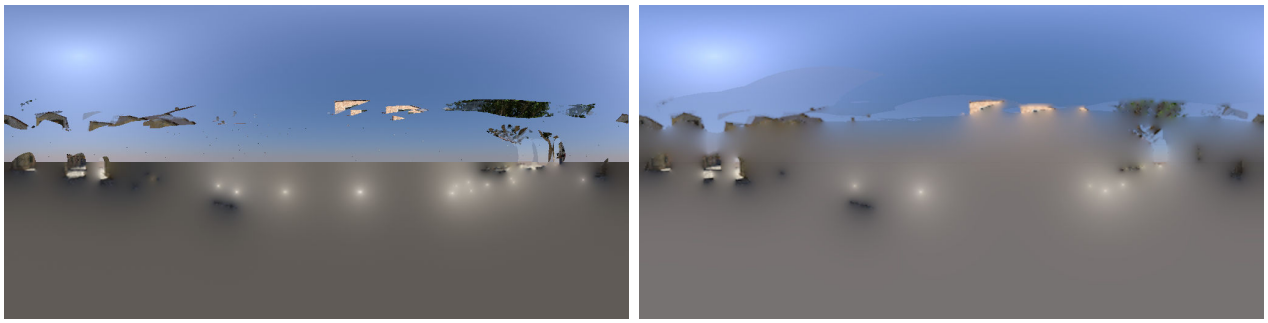
Figure 5.24: Overview of the reconstructed geometry from the synthesized images.



Input envmap. Our environment map.

Figure 5.25: Reference color jet map used to represent the mean squared error per pixel for reflectance Fig. 5.30 and environment Fig. 5.31.

We purchased a model of a scene which has a similar appearance to the real environments we target, with realistic textures for the building, densely foliaged trees and we used a physically-based sky model [Preetham *et al.*, 1999]. We used an in-house path-tracer to render 44 images (Fig. 5.23), which we took as input for our complete pipeline including the multi view stereo algorithm (Fig. 5.24). Multi-view stereo has difficulty with synthetic models and textures, and the quality of the reconstruction is poor, as can be seen in the inset; large portions of the tree are not well reconstructed and the overall geometry is coarse and approximate. We also rendered the corresponding layers of reflectance and shading for



Our environment map. [Laffont, 2012]

Figure 5.26: Overview of our synthesized environment on the left and the method described in [Laffont, 2012] on the right.

quantitative comparison, respectively shown in Fig. 5.30 and Fig. 5.31.

LMSE=0.067 GMSE=0.104 | LMSE=0.070 GMSE=0.109

LMSE=0,044 GMSE=0,066 | LMSE=0,044 GMSE=0,065

(a) Input image | (b) Reconstructed proxy | (c) Our reflectance | (d) Ground truth reflectance | (e) [Laffont *et al.*, 2013]

Figure 5.27: Comparison between our method, [Laffont *et al.*, 2013] and ground truth reflectance rendered from a synthetic scene. Our method produces a few strong yet localized errors due to mis-classification of small regions in the shadow of the tree. In contrast, [Laffont *et al.*, 2013] exhibits a low yet extended deviation from ground truth in the shadow region. The two methods are quantitatively similar according to the LMSE and GMSE error metrics.

Figure 5.27 provides a visual and quantitative comparison of our reflectance against ground-truth and the result of Laffont *et al.* [2013]. We selected the parameters of [Laffont *et al.*, 2013] that produce the best decomposition. The two methods yield results of similar quality as measured by the local mean squared error **LMSE** and the global mean squared error **LMSE** error metric [Grosse *et al.*, 2009]. Since the intrinsic decomposition problem is still largely unsolved, **LMSE** measurement allows us to weight the per pixel local mean squared error by the mean over a window size. Both **LMSE** and **LMSE** evaluate the quality of an intrinsic decomposition method. However, close inspection reveals that most

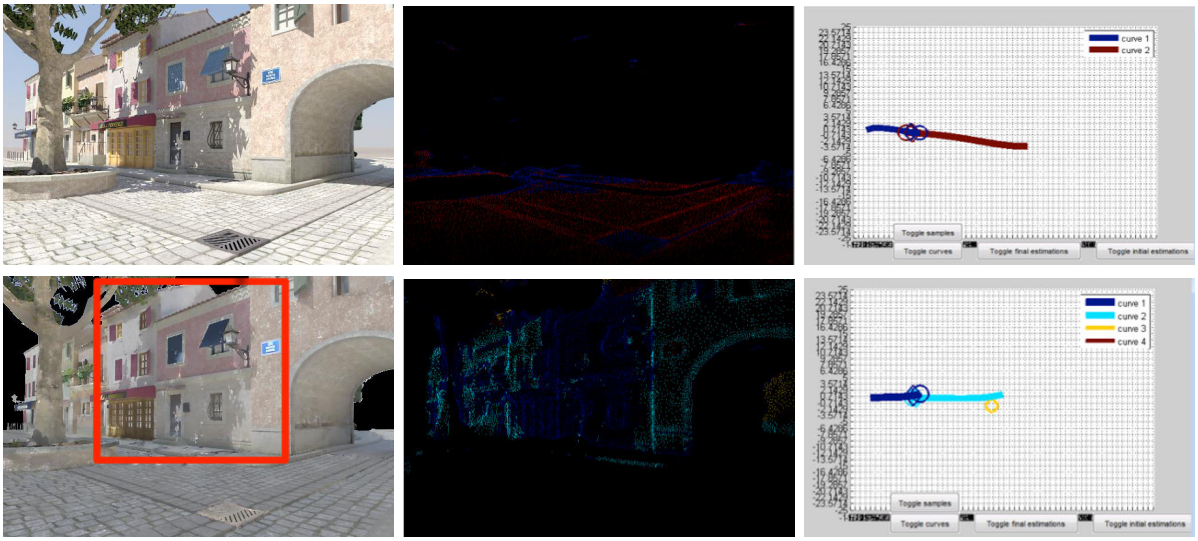


Figure 5.28: Intersection in LAB space of reflectance candidate curves per normal orientation for [Laffont *et al.*, 2013].

Input image | Our reflectance | Our propagated visibility

Figure 5.29: Highlight classification error in our model.

of our error is due to mis-classification of small shadow regions, which yields strong yet localized deviation from ground-truth, while [Laffont *et al.*, 2013] fails to completely remove the shadow of the tree on the wall, which yields a low yet extended erroneous region (Fig. 5.27, top, far right). This different type of error is due to the fact that the method of Laffont et al. does not explicitly estimate binary visibility and does not refine the estimation of environment shading as shown in Fig. 5.28; our approach yields results more suitable for relighting to be consistent.

Figure 5.30 visualizes our error on reflectance and Figure 5.31 environment shading. This visualization reveals that a significant part of our error is due to the approximate environment shading, especially in areas where this component dominates sun shading. On the one hand, we refine the environment shading near shadow boundaries which allows relighting and on the other hand it is also the main limitation of our refinement step since we cannot refine it for regions far from the shadow boundaries.

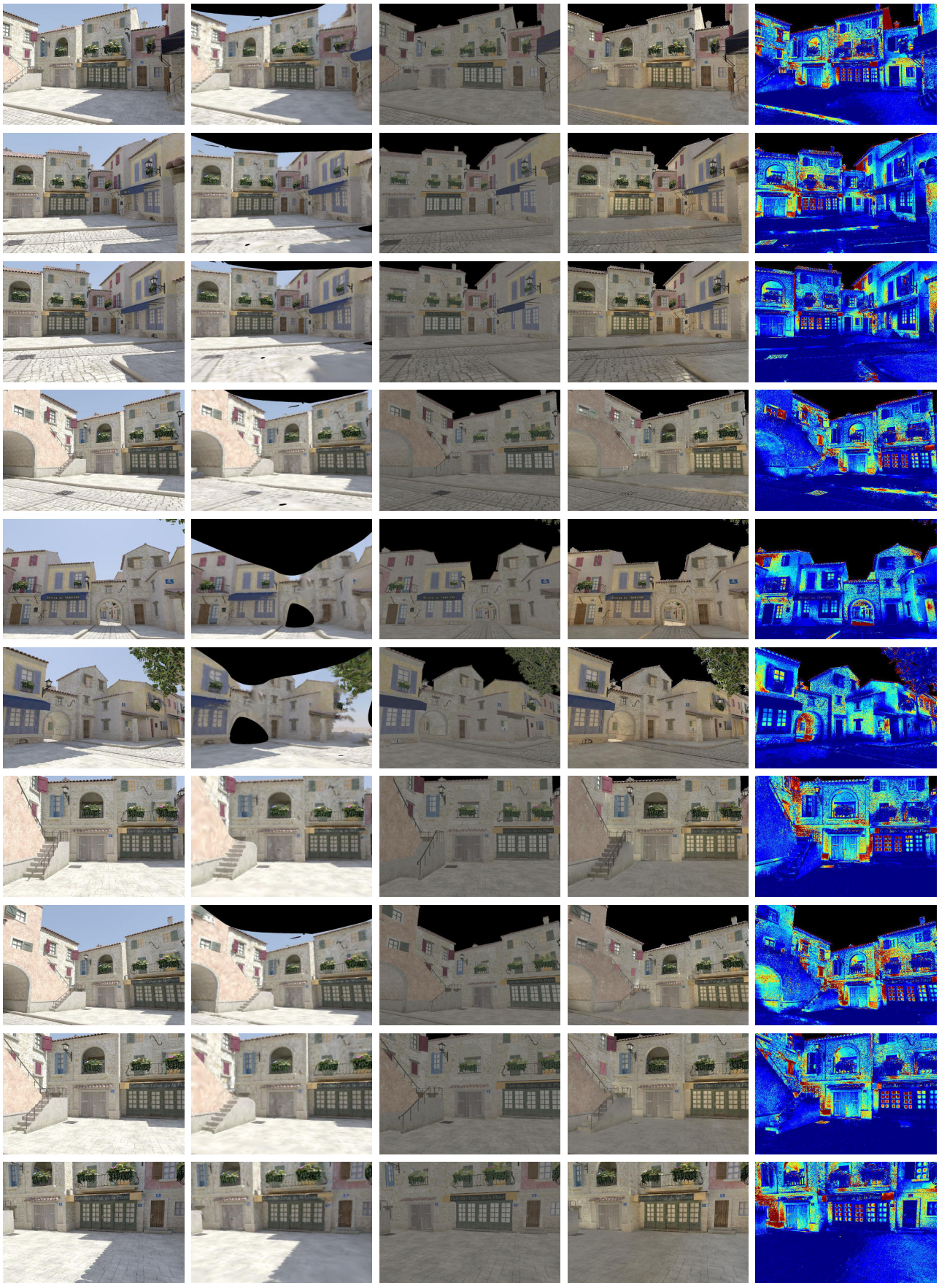


Figure 5.30: Ground truth GT results and mean squared error **MSE** for reflectance.

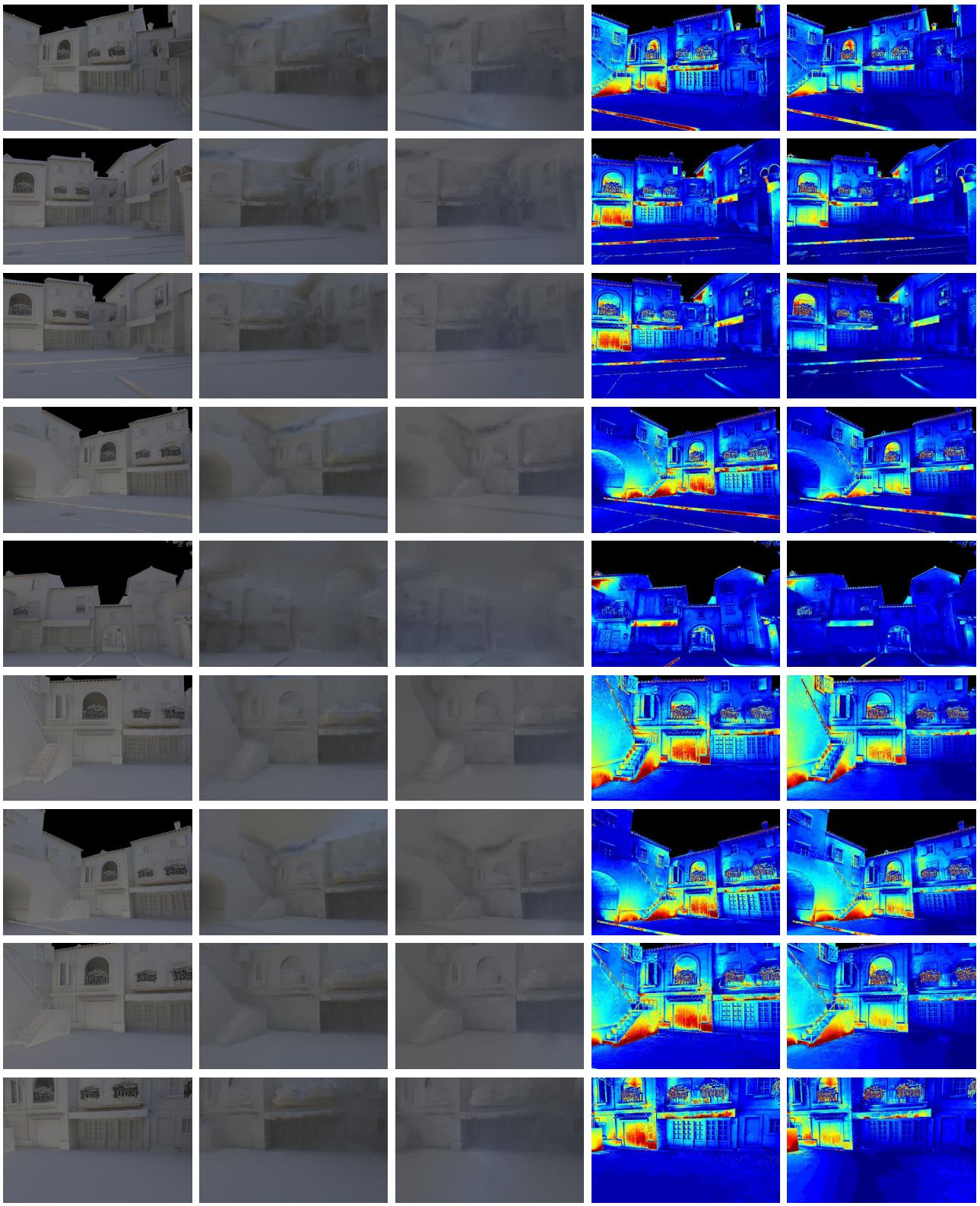


Figure 5.31: Ground truth GT results and mean squared error **MSE** for the environment Env and refined Env.

5.7 Ground Truth relighting comparison

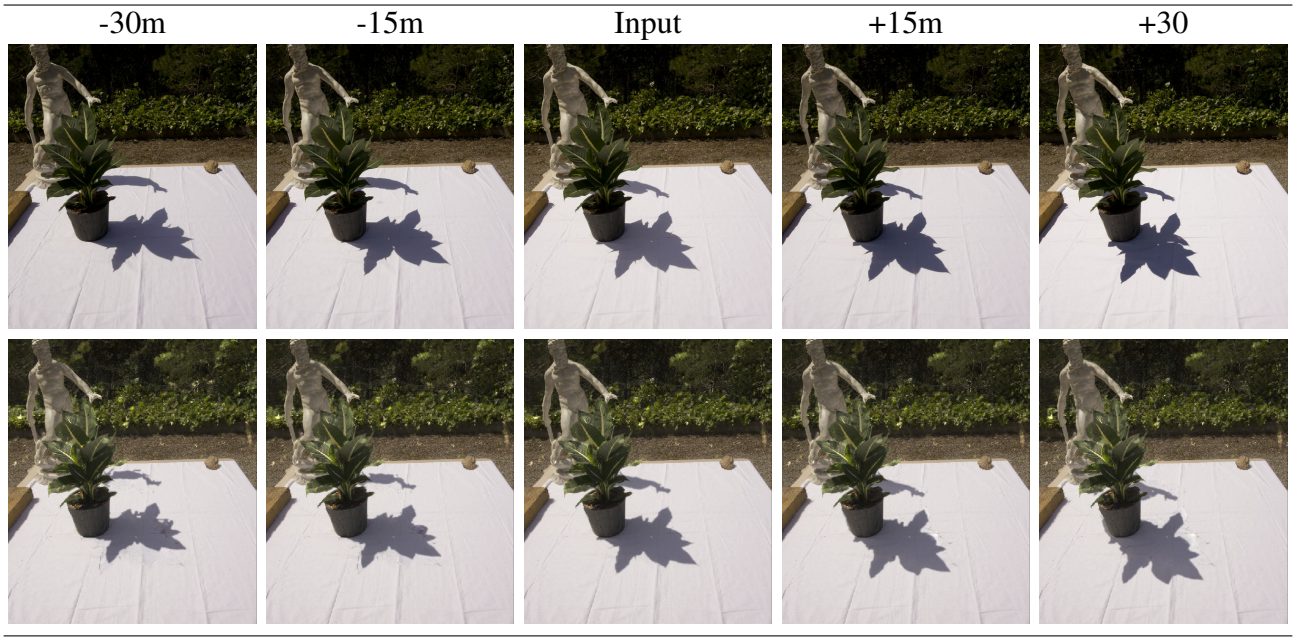


Figure 5.32: Above: real photographs taken at different times than those used for the algorithm. Below: relit images using our algorithm.

We captured several lighting conditions for the Plant scene to allow a ground truth comparison. We only used multi-view capture of the central image (i.e., a single lighting condition) for all intrinsic decomposition and relighting computations. We show the results in Fig. 5.32. We can see that the cast shadow becomes more approximate as we move away from the time of capture used by our algorithm, but the overall appearance is plausible. A slight residue of the original penumbra remains visible in the reflectance, which is due to the non-diffuse nature of the white tablecloth we placed on the table. Since the camera is close to the glossy lobe of this surface, our assumption of diffuse reflectance reaches its limits and our refinement step is not sufficient to fully correct for the remaining errors. Note also that our synthetic shadows have the same color as the shadow in the input image because they are computed from an estimate of the same sun color and sky model. In reality the appearance of the sky changed over time, which explains why the real shadow is darker in some pictures.

5.8 Conclusion

The extensive evaluations of our intrinsic image and relighting algorithms presented in this chapter allowed us to better understand their strengths and weaknesses. The ground truth comparison as well as our comparisons to previous algorithms and our evaluation of robustness to reconstruction quality have demonstrated that our algorithms performs well in many challenging conditions. Evidently, our

algorithms do not always succeed, and are only a first step in the quest to solve these very hard and challenging problems. The limitations of our algorithms analysed with the evaluation presented here served as a basis for our proposals for future work, discussed in Chapter 7.

Chapter 6

Estimating Image Based Bidirectional Reflectance Functions

6.1 Motivation

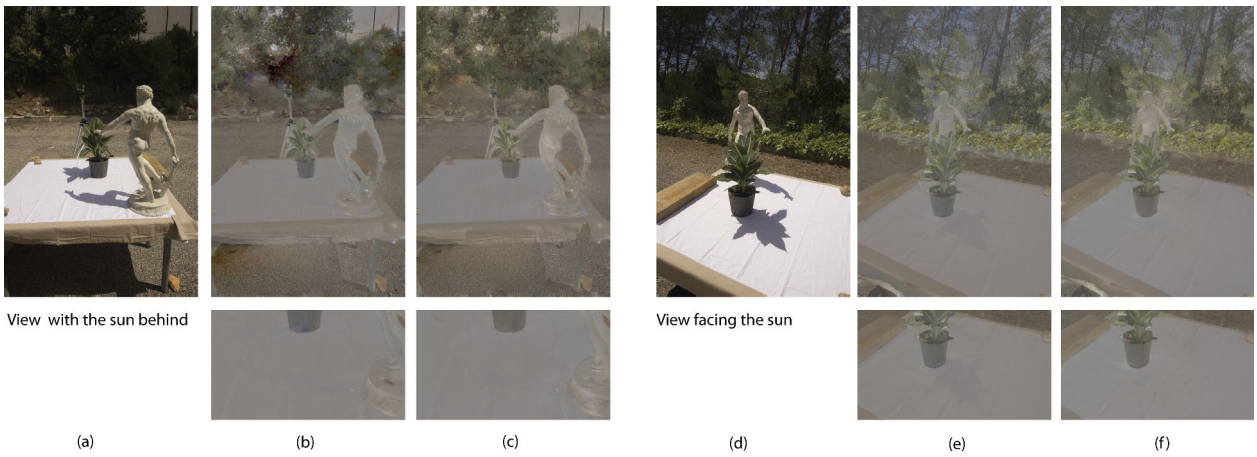


Figure 6.1: The captured image (a) is taken with the sun behind which implies no specular effect, (b) shows the obtained reflectance using the sun calibration, the shadow classifier and propagation without the local correction on boundaries, (c). Note there is no major difference between (b) and (c). However, for a view facing the sun which maximizes the specular effect(d), the local correction near boundaries compensates this error (e), (f).

In Chapter 3, we demonstrate that we can recover the reflectance of a surface from inaccurate geometry allowing us to manipulate the lighting condition and also the shadow in Chapter 4. The restruction to Lambertian reflectance leads to artifacts that are hidden by the correction per image of the environment lighting. The captured image(a) shown in Fig. 6.1 is taken with the sun behind which implies minimal specular effect. Image (b) shows the obtained reflectance using the sun calibration, the shadow classifier, and the image driven lighting propagation without the local correction on bound-

aries, (c). Note that there is no difference between (b) and (c). However in the case of a view acquired facing the sun (d), the specular component is near its maximum and in this case our local environment refinement step compensates for the error due to specular effect more than the mis-estimation of lighting condition as shown in (a). Tackling this shortcoming requires to model more complex BRDFs per point than the Lambertian model used in Chapter 3.

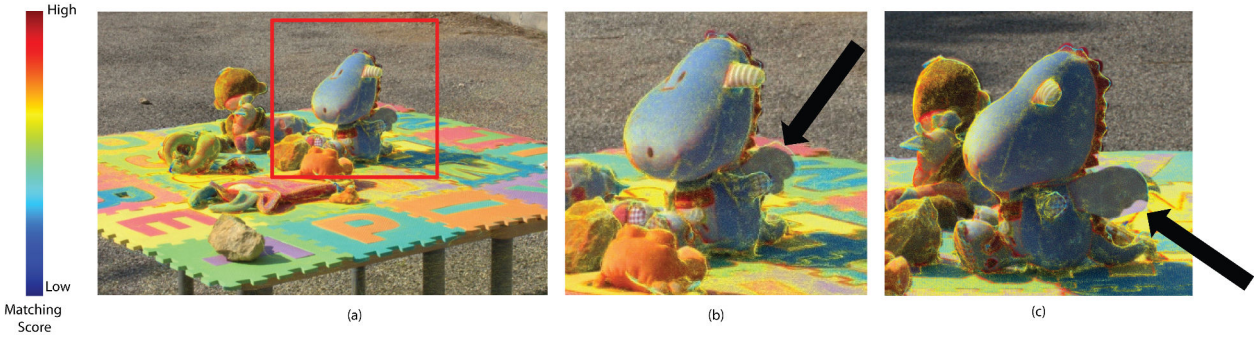


Figure 6.2: Matching confidence score splatted in one input view(a). (b) and (c) show local inaccuracies and projection errors inherent to a multi view reconstruction algorithm in adjacent views of (a).

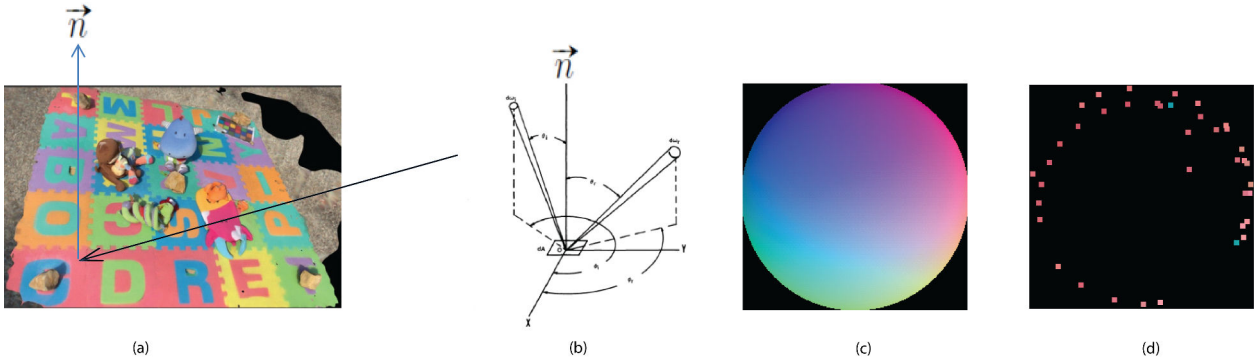


Figure 6.3: (a) Colored vizualization of the proxy using the median of radiance for a point of normal $\mathbf{n}$. (b) Radiance hemisphere in 3D. (c) Normal hemisphere view from top in 2D. (d)Top view of the distribution of radiance samples in the frame (c). Each sample is splatted on this hemisphere using the color from the corresponding input cameras.

Retrieving BRDFs or spatially varying BRDFs from images is an active field. The thesis of Steve Marchner [Marschner, 1998] largely inspired subsequent methods which use known surfaces such as a sphere [Matusik *et al.*, 2003], [Romeiro *et al.*, 2008], [Romeiro and Zickler, 2010a] or very precise and dense captured geometry [Yu *et al.*, 1999], [Lensch *et al.*, 2001], [Debevec *et al.*, 2004]. However, the context of geometry acquired using multi view stereo method shown in Fig. 6.2 introduces two new challenges, noise and sparsity, Fig. 6.3. The main goal of this study is to address these issues.

- Noise is the result of projection error, since a reconstructed 3D point may not be properly project in all registered views as shown in Fig. 6.3, (d).

- Sparsity comes from the small number of images used to reconstruct a scene, indeed in our context we typically use 40-80 images as input for a complex scene involving many objects whereas previous methods can use the same number [Lensch *et al.*, 2001] or hundreds [Debevec *et al.*, 2004] only to acquire one object. In our case, a point cannot contain more radiance samples that the input views used for reconstruction which directly implies a very sparse sampling of the outgoing radiance hemisphere Fig. 6.3, (d).

6.2 Our approach

Our context is strictly different than previous methods since we performed the capture in an uncontrolled lighting environment without a moving light source to generate more samples on the outgoing radiance hemisphere shown in Fig. 6.3. Moreover, we do not assume knowledge about the shape of the object and we do not consider only one object in our scene. Our geometry is inaccurate and many projection errors can be identified. We also do not use HDR images, just the RAW file coming from the DSLR camera.

Recall the reflectance expression Section 2.3.4, Eq 2.15, in the cases, we assume an ideal diffuse reflection function known as Lambertian we can ignore the incoming illumination direction to reflect outgoing radiance and the obtained BRDF becomes a constant $f_{r,d}$:

$$L_r(x, \overrightarrow{w_r}) = \int_{\omega_i} f_{r,d}(x, w_i, w_r) L_i d\Omega_i = f_{r,d}(x) \int_{\omega_i} L_i d\Omega_i = f_{r,d}(x) E_i(x) \tag{6.1}$$

As explained previously, a Lambertian BRDF model is not representative of the surface appearance of an object. In previous work requiring image based BRDF estimation, the Lafortune BRDF [Lafortune *et al.*, 1997] is widely used since this model can cover a wide range of materials which can be captured. In the case of physically based shading, considering a non-diffuse energy conserving BRDF, direct and indirect lighting, the rendering equation can be divided into sub-components according to a BRDF including two components; a diffuse term and a specular term as described in [Pixar Animation Studios and Villemin]:

$$L_r(x, \overrightarrow{w_r}) = \int_{\omega_i} f_{r,d}(x, \overrightarrow{w_i}, \overrightarrow{w_r}) L(x, \overrightarrow{w_i}) d\Omega_i \tag{6.2}$$

$$L_r(x, \overrightarrow{w_r}) = \int_{\omega_i} f_{r,d}(x, \overrightarrow{w_i}, \overrightarrow{w_r}) (L_{dir}(x, w_i) + L_{ind}(x, \overrightarrow{w_i})) d\Omega_i \tag{6.3}$$

$$L_r(x, \overrightarrow{w_r}) = \int_{\omega_i} (f_{diff}(x, \overrightarrow{w_i}, \overrightarrow{w_r}) + f_{spec}(x, \overrightarrow{w_i}, \overrightarrow{w_r}))(L_{dir}(x, \overrightarrow{w_i}) + L_{ind}(x, \overrightarrow{w_i}))d\Omega_i \tag{6.4}$$

$$\begin{aligned} L_r(x, \overrightarrow{w_r}) = &\int_{\omega_i} f_{diff}(x, \overrightarrow{w_i}, \overrightarrow{w_r}))L_{dir}(x, \overrightarrow{w_i})d\Omega_i \\ &+ \int_{\omega_i} f_{spec}(x, \overrightarrow{w_i}, \overrightarrow{w_r})L_{dir}(x, \overrightarrow{w_i}))d\Omega_i \\ &+ \int_{\omega_i} f_{diff}(x, \overrightarrow{w_i}, \overrightarrow{w_r}))L_{ind}(x, \overrightarrow{w_i})d\Omega_i \\ &+ \int_{\omega_i} f_{spec}(x, \overrightarrow{w_i}, \overrightarrow{w_r})L_{ind}(x, \overrightarrow{w_i})d\Omega_i \end{aligned} \tag{6.5}$$

$$L_{rspecularindirect}(x, \overrightarrow{w_r}) = \int_{\omega_i} f_{spec}(x, \overrightarrow{w_i}, \overrightarrow{w_r})L_{ind}(x, \overrightarrow{w_i})d\Omega_i \tag{6.6}$$

To retrieve both specular and diffuse components from the BRDF, we are going to ignore the specular component due to indirect paths described in Eq 6.6. The main reason to ignore this term is because of its complexity of evaluation. Indeed to be modelled properly, we first need to know the BRDF for all points of the scene which would require to bootstrap over the optimization to refine the solution on a iterative way. This is not the goal of this study. The second reason is that our capture sampling strategy is really sparse. Capturing radiance samples combining diffuse and specular terms due to the main light source, the sun, is already very hard. Recall that we capture only 40-80 images for a full scene.

Indirect illumination is then evaluated as described in Chapter 3, by estimating the irradiance at each 3D point. For each casted ray, we use a linear interpolation over the 3 vertices of the intersected triangle and their outgoing radiance obtained by taking the median of the collected radiance from the input views. Of course, interpolation between 3D points is more or less accurate according to the density of the 3D reconstructed point but no existing metrics exist to estimate the reconstruction confidence of a point. In such context, the matching score used for the reconstruction is not suitable as shown in Fig. 6.2. The direct consequence of these issues is that we give an equal confidence to all the integrated samples.

For this image based BRDF retrieval, we use the following image formation model:

$$\begin{aligned} L_r(x, \overrightarrow{w_r}) = & \int_{\omega_i} f_{diff}(x, \overrightarrow{w_i}, \overrightarrow{w_r}))L_{dir}(x, \overrightarrow{w_i})d\Omega_i \\ & + \int_{\omega_i} f_{spec}(x, \overrightarrow{w_i}, \overrightarrow{w_r})L_{dir}(x, \overrightarrow{w_i}))d\Omega_i \\ & + \int_{\omega_i} f_{diff}(x, \overrightarrow{w_i}, \overrightarrow{w_r}))L_{ind}(x, \overrightarrow{w_i})d\Omega_i \end{aligned} \tag{6.7}$$

In such an undetermined problem, a piecewise Linear interpolation might have been an option. However detecting outliers due to the inaccuracy of the geometry and projection error on the radiance hemisphere would involve running an optimization to discard them. Moreover, such interpolation can be used to re-render interpolated views but does not bring extra information either to evaluate complex light transport paths such as indirect specular or to manipulate BRDF models. In the following, we introduce an optimization that is only valid for an outdoor scenes with one main light source which is the sun. The main motivation of this study is to dress the list of issues by starting with a smaller problem than the indoor scene scenario. In the context of an outdoor scene, direct lighting only comes from the sun and indirect lighting can be roughly estimated by using the method described in Chapter 3. In terms of light transport simulation of outdoor scene, the sun is considered like a light source at infinity which cannot be applied to indoor lights. In the case of an indoor scene, light sources would have to be more carefully characterized in term of fall off and intensity. Given signal processing theory, the Nyquist-Shannon theorem implies that the outdoor scene problem is already undetermined since the sparsity per points of the outgoing radiance cannot lead to an optimal solution. Indoor scenes are not tackled because multiple light source characterization to retrieve BRDF is even more largely undetermined.

The method of [Lensch *et al.*, 2001] develops an approach of sharing samples of the same material to complete the outgoing radiance hemisphere on clusters of points in a controlled environment and a single object. In a second step, spatially BRDFs are estimated by expressing the shading model for each point as a linear combination of a collection of mutated BRDFs base obtained from this sharing strategy. However, [Matusik *et al.*, 2003], [Bonneel *et al.*, 2011] demonstrate that direct linear interpolation of BRDF leads to strange behaviour in the specular component and non-plausible BRDFs. Note that this shortcoming is adressed with manifold-based interpolation in [Matusik *et al.*, 2003], [Dong *et al.*, 2010]. However as described in Chapter 2, building such a basis requires BRDF samples. For this reason, these methods are not shown.

6.3 Lafortune BRDF

This BRDF model is a combination of cosine-lobes centered around different axes and a traditional diffuse component.

$$f_{Laf}(x, \overrightarrow{w_i}, \overrightarrow{w_r}) = \rho_d + \sum_k \rho_{s,k} * s_k(\overrightarrow{w_i}, \overrightarrow{w_r}) \tag{6.8}$$

ρ_d is the diffuse reflectance, $\rho_s * s$ is a lobe component. The lobe s is expressed with a standard matrix notation. In our context, we only address isotropic reflection so we use only a diagonal matrix C as shown in Eq.6.9 and the lobe can be expressed as a dot product Eq. 6.10; modelling anisotropic reflection involves the used of a 3×3 matrix. As C defines the orientation of the lobe, the C coefficients can shear the lobe in different ways to represent different surface-scattering behavior.

$$s(\overrightarrow{w_i}, \overrightarrow{w_r}) = \left(\begin{bmatrix} w_{r,x} \\ w_{r,y} \\ w_{r,z} \end{bmatrix}^T \begin{bmatrix} C_x & & \\ & C_y & \\ & & C_z \end{bmatrix} \begin{bmatrix} w_{i,x} \\ w_{i,y} \\ w_{i,z} \end{bmatrix} \right)^{n_k} \tag{6.9}$$

$$s(\overrightarrow{w_i}, \overrightarrow{w_r}) = (C_x w_{i,x} w_{r,x} + C_y w_{i,y} w_{r,y} + C_z w_{i,z} w_{r,z})^{n_k} \tag{6.10}$$

This representation already offers plenty of variety for each lobe, and the complete BRDF can model diffuse reflection, retro-reflection and specular reflection. Another key aspect of the isotropic case is the reduction of one unknown to solve since to guarantee isotropy $C_x = C_y$.

In the case the matrix C is defined as $C_x = -1, C_y = -1, C_z = 1$; the incoming light $\overrightarrow{w_i}$ is reflected about the normal of the point x, and the expressed lobe is specular as in Phong BRDF. Retro-reflection is modeled by the lobe when $Cx = 1$, $Cy = 1$, $Cz = 1$ since the identity matrix means in this case that the incoming light $w_{i,x}$ is reflected in the same direction, $\overrightarrow{w_i} = \overrightarrow{w_r}$.

6.4 Constrained Non Linear Fitting

To retrieve the BRDF, we will solve:

$$\begin{aligned} L_r(x, \overrightarrow{w_r}) = & \int_{\omega_i} f_{diff}(x, \overrightarrow{w_i}, \overrightarrow{w_r})) L_{dir}(x, \overrightarrow{w_i}) d\Omega_i \\ & + \int_{\omega_i} f_{spec}(x, \overrightarrow{w_i}, \overrightarrow{w_r}) L_{dir}(x, \overrightarrow{w_i})) d\Omega_i \\ & + \int_{\omega_i} f_{diff}(x, \overrightarrow{w_i}, \overrightarrow{w_r})) L_{ind}(x, \overrightarrow{w_i}) d\Omega_i \end{aligned} \tag{6.11}$$

$$arg\min_{\mathbf{x}} \sum_{i}^{m} |L_r(x, \overrightarrow{w_r})_{obs_m} - L_r(x, \overrightarrow{w_r}, f_{Laf})_{est_m}|^2)$$

with constraints on: diffuse parameter, $0 \leq \rho_d \leq 1$ (6.12)

specular parameters and isotropy, $0 \leq \rho_s \leq 1, C_x = C_y, 1 \leq n \leq 100$

energy conservation, $0 \leq \rho_d + \rho_s \leq 1$

Levenberg Marquadt is a very popular optimization scheme to solve non-linear least squares problems. However, its implementation can make strong difference in the quality of the result. So far, we found that the implementation provided by [Agarwal *et al.*] in **Ceres-solver** gives the best results. In the context of BRDF retrieval, we estimate the square difference for each point between their input radiance samples collected from multiple views and the rendering of their estimated BRDF for the parameters ρ_d, ρ_s and n under our estimated lighting condition. We follow the strategy to fit one BRDF per color channel RGB, and to bound each of theirs parameters to prevent the optimization from converging to undesired values and to make sure the BRDF obtained conserves energy. The problem is initialized by using the median of the radiance for the diffuse component and the specular part is initialized as a mean BRDF.

Since we know the sun direction, we can constrain the optimization on the BRDF to take into account the sun direction, as well as the sun color and the irradiance. Indeed, we expect a specular component to be observed in the mirror direction of the sun which is model by creating a lobe in this direction.

6.5 Results

To evaluate the quality of the fitting model we render each point from the 3D model using our estimated BRDFs under our lighting conditions and the normal provided by the 3D reconstruction. For display reason, we propagate these estimation using an image driven method [Levin *et al.*, 2008a]. Then we compute the L_1norm between the corresponding input view image and the fitting model. Each color channel is directly added to the error measurement which uses the color jetmap of matlab between 0 and 255. Results are shown in Fig. 6.4, 6.5.

Note that local correction of the misestimation of indirect illumination was not performed and the re-rendering can suffer from this approximation. However the non linear optimization should have compensated for this lack during the optimization.

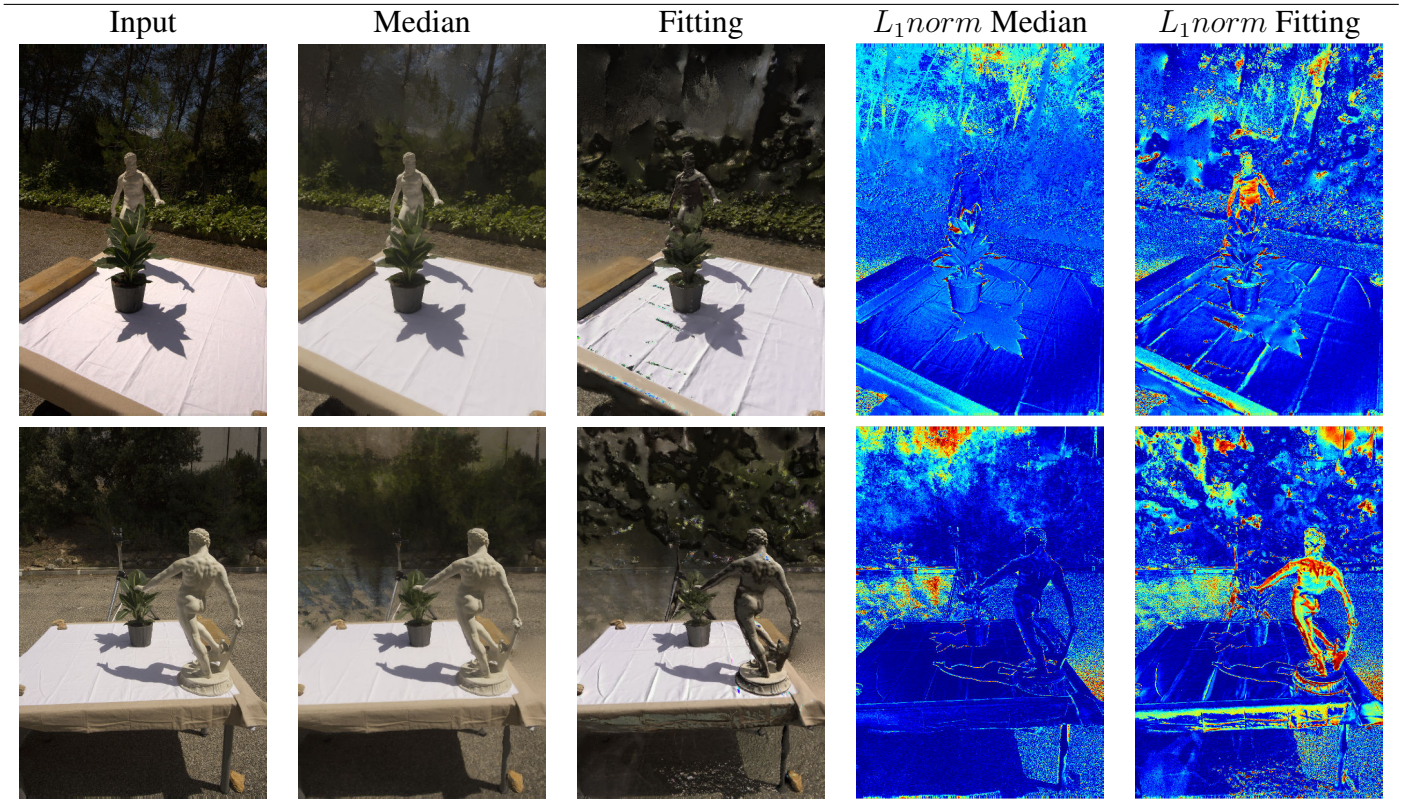


Figure 6.4: Statue and plant scene for the views 1 and 21. Input column shows the input image, Median and Fitting respectively shows the rendering of our estimation by taking the median or by using BRDF fitting. L_1norm columns show the difference between the input image and the rendering.

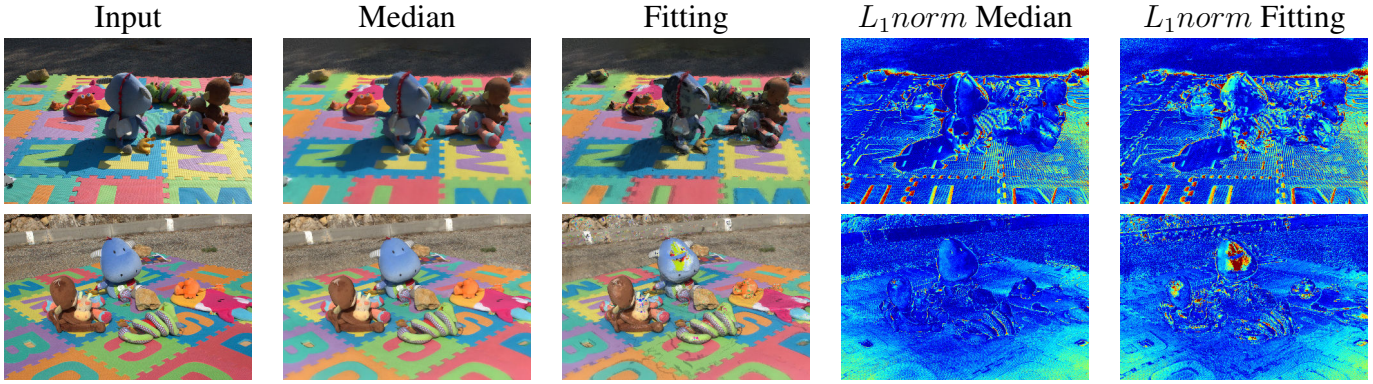


Figure 6.5: Toys scene for the views 23 and 32. Input column shows the input image, Median and Fitting respectively shows the rendering of our estimation by taking the median or by using BRDF fitting. L_1norm columns show the difference between the input image and the rendering.

In both cases, this study highlights that the outgoing radiance of the hemisphere can reasonably be well approximated by computing its median. We can note that the fitting of the statue in marble Fig. 6.4 fails independently of the view selected to render. First, despite ignoring the specular light due to the indirect component in Eq.6.6, in this case the indirect light coming from the table to the statue is not

evaluated but in the case of the pot on the table and the wood the L_1norm is reasonable even in the case they are lighted only by the indirect light from the table. We know that complex materials such as marble cannot be modelled efficiently by a Lafortune BRDF but we were expecting a better fitting. In Fig. 6.5, the fitting on the head of the main teddy bear in blue is also unstable.

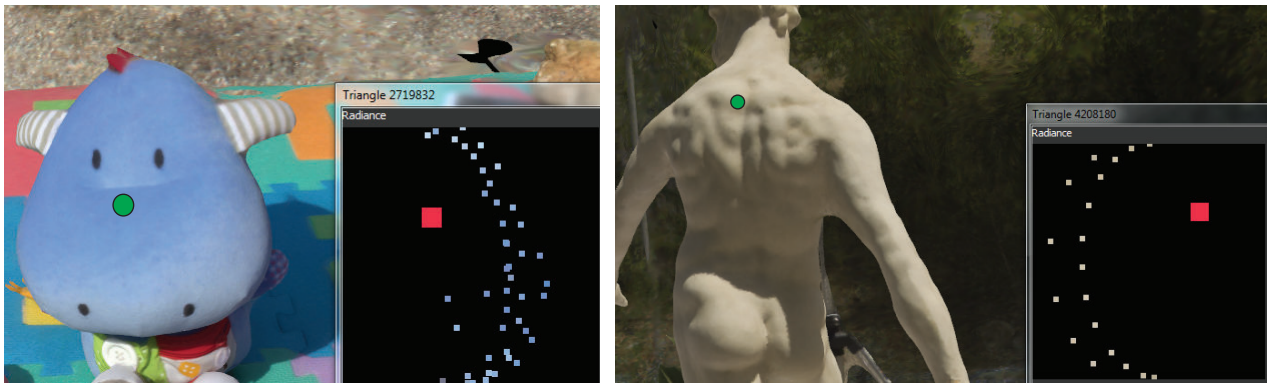


Figure 6.6: The green spot indicates the clicking vertex. In both cases, the distribution of samples only cover a really small portion of the hemisphere.

By looking closely at the unfitted regions in Fig. 6.3, the capture of the outgoing radiance hemisphere does not really covered the azimuth angle of 360 degrees. However, for a small differential angle on the azimuth axis, the zenith angle is reasonably sampled. A future study should then focus on similar objects as the statue and teddy bear to evaluate if these unfitted regions are really due to the material complexity. We can imagine that the method of [Lensch *et al.*, 2001] would give successful result on these scenes, however with a long term goal of manipulating the BRDFs of the scene, this would involve the manipulation of a set of basis per points instead of manipulating a full region which is non-intuitive for end users.

6.6 Conclusion

In this chapter, we studied the problems related to the limitation of our reflectance model to simple Lambertian in the previous chapters. After providing more detail on the issues involved, related to very sparse and inaccurate data, we conducted a first experiment to estimate a Lafortune model based on this sparse sampling. Even though the first results are quite limited, we believe that this is a very interesting direction for future work, both for the simple extraction of materials and for more accurate and effective intrinsic image algorithms.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

In this thesis we presented the first method to allow automated intrinsic image decomposition for multi-view outdoor scenes with cast shadows, providing reflectance, shading and cast shadow layers at a quality level which is suitable for relighting. We were thus able to introduce multi-view relighting, and demonstrate its utility for Image Based Rendering with changing illumination. Apart from IBR, other applications of multi-view relighting are possible, for example in compositing for post-production, where lighting changes often involve a significant amount of manual work. By allowing multi-view relighting, our solution takes an important step in making image-based methods a viable alternative for digital content creation. The most important insight obtained in this thesis is that progressive refinement built upon constrained pairs from rough estimation can lead to high quality estimation of the reflectance. However, building these constraints is a real challenge especially in regions where no boundaries can be estimated to reduce the search space to place them. Our ground truth evaluation showed that our local correction is limited to shadow boundaries and can actually increase the error in some regions. In term of contributions, this thesis introduced an automatic environment lighting estimation, a shadow classification based on inverse rendering, a novel intrinsic decomposition method, a relighting algorithm, a ground truth evaluation and a study of the image based BRDF measurement in the context of multi view stereo 3D reconstruction.

7.2 Research impact and deployment

The results of this thesis are being used in the CR Play project `www.cr-play.eu` which targets the development of image-based rendering methods as a tool for content generation in games. Such systems require manipulating texture and lighting conditions of a multi view captured scene. Our multi

view intrinsic decomposition algorithm was also transferred to Autodesk. Autodesk was also a partner in this project and gave us access to their multi-view stereo reconstruction pipeline through 123D Catch.

7.3 Future work

Our approach opens up many possibilities for future work which focus on several axes: targeted scenes, user interaction, perception, segmentation and reconstruction.

Scenes. Currently, our approach assumes outdoors scenes with sunlight and well-defined cast shadows. In this thesis we assume our environment map can be completed by inpainting for the ground under the horizon and by fitting a sky model above the horizon. For scenes with overcast sky, the problem can appear simpler, since the variation between shadow and light is much smoother. Precise determination of shadow boundaries is thus unnecessary. However, our approach must be extended to handle such soft boundaries, possibly with a new soft shadow classifier approach. The importance of local correction was demonstrated in the evaluation section and performing a similar process will be a challenge in the case of soft boundaries.

Material characterization. Current intrinsic decomposition methods assume only Lambertian materials. One conclusion of this thesis is the need to propose a more complete image formation model that incorporates non-diffuse behavior. This is an exciting fundamental research direction which requires a completely new approach to intrinsic image decomposition. Material characterization is also a critical aspect to target indoor scenes which highlights two mains challenges to be solved at the same time: characterizing light sources and materials. By opposition to outdoor scenes where only the sun is considered at infinity, a lamp has a non-uniform emittance model and a fall-off which need to be taken into account. However to characterize such complex lights, we can easily imagine a regression over a pair of points sharing the same material and different lighting conditions but the complexity of materials (e.g glossy) will make such characterization hard. As shown in the Chapter 6, the inaccuracy and sparsity of the acquired radiance samples in the context of a nonlinear optimization introduces new challenges. Bell *et al.* [2014] and Karsch *et al.* [2014b] tackle indoor scenes with two different approaches respectively based on human judgements and on learning techniques which could be unified in the same framework. Moreover both of these approaches focus on single images and could take advantage of a multi view stereo acquisition system.

Dynamic scenes and video. Dynamic scenes and video based intrinsic decomposition is a new axis of research limited to texture edition, grayscale shading and basic mixing of two scenes until now. Two different approaches favoring either an automatic method [Ye *et al.*, 2014] or user feedback [Bonneel *et al.*, 2014] perform an intrinsic decomposition to allow texture edition and provide temporal coherence of edition. However, both methods are limited to processing a sequence without strong motion of the camera or view angle changes during the sequence, and basic lighting conditions. Lighting condition manipulation is not shown, and the case of complex lighting conditions with multiple light sources in a street cannot be treated. It might be interesting to consider the case of a sequence recorded from multiple views during a single event where no reshooting is allowed. In such cases material appearance will play a key role to exploit temporal coherence over time and angle views.

User interaction and Feedback. Our method takes about 2 hours to process a scene with 80 input images. An user interface could be imagined to correct some classification issues but the local refinement step will still take about 1 minute per image with optimized code, which makes the method impractical for an artist. We think this aspect is also neglected in several other method using a single lighting condition. In our comparisons Section 5.4, the method required:

- Barron and Malik [2013a] about 2 days.
- Chen and Koltun [2013] 20 minutes.
- Laffont *et al.* [2013] about 1 day of processing and several hours to select the parameters for one input view.

However, one recent method [Bonneel *et al.*, 2014] performs the intrinsic decomposition in half a second to allow the user to fix some local errors as shown in Fig.7.1, in this context some manual editing is possible but the approach described is limited to a grayscale lighting layer and texture edition. The method still requires up to 20 minutes of manual strokes to achieve the desired result on a video.

Figure 7.1: Input image from the video sequence. Purple regions impose a constant shading with the method of [Bonneel *et al.*, 2014] (a), temporal coherence propagation with the method of [Bonneel *et al.*, 2014] (b), the method of [Ye *et al.*, 2014] (c).

Perception. In this thesis, we did not evaluate the perceived quality of our scene manipulation with a perception study. However it is important to keep in mind that it is not clear that a perfect inverse rendering algorithm is required to manipulate a scene especially for texture edition, which suggests that a perception study should be run to validate a manipulated scene and understand what disturbs a viewer the most. Ideally we would like to obtain a perfect intrinsic decomposition to manipulate lighting conditions, reflectance properties such as BRDF, texture or mix several captured objects in the same scene. This highlights that a complex algorithm is not necessarily required for all scene manipulations.

Bonneel *et al.* [2014] described an interactive intrinsic decomposition method for video which produces an inaccurate intrinsic decomposition, but allows coherent texture edition over a video and to blur shadow. However, some strong shadow residuals remain in the reflectance layer which forbade any motion of the shadow without producing disturbing artefacts as shown in Fig 7.2.

The user study of Karsch *et al.* [2014b] shows as well that many rough CG insertions not perturb the final user despite having inconsistent lighting between the scene and the mixed object. Note that this method roughly estimates reflectance in order to obtain the most likely lighting condition to provide convincing results.

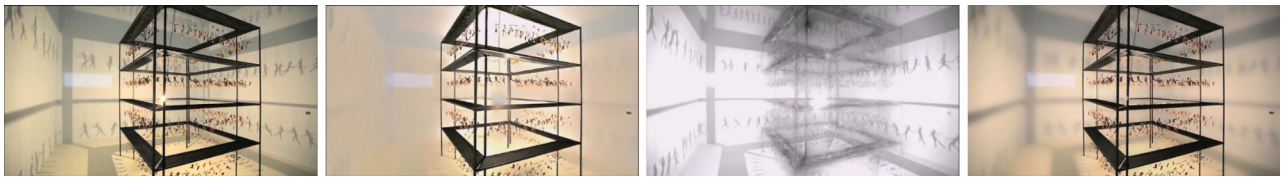


Figure 7.2: Įnput image from the video sequence(a), reflectance(b), grayscale shading (c), Shadow softened (d). From the method of [Bonneel *et al.*, 2014].

Segmentation and reconstruction. The shadow classifier we presented is incomplete for two reasons: ideally we would like to connect each node to each other to propagate the most likely belief instead of selecting a subset which behaves like a filter. As described in Chapter 3, irradiance from sky and indirect lighting are computed for each 3D point, but the estimation of the visibility label is not shared across views but rather performed per view. This lack of shared information between multiple views is also responsible for the misclassification of some isolated clusters. Another aspect inherent to segmentation techniques is the difficulty to handle thin objects which cannot be clustered properly. A graph representation which connects all clusters to each other can easily lead to unstable solutions since some energy terms are non-sub modular requiring to solve the graph using solvers such as loopy belief propagation or TRW-S which are not known for their stability but provide a solution to this NP-hard problem. Concerning the representation of a graph built on multiple 2D images, currently there is no ideal way to transfer information from an inaccurate 3D reconstructed point to all visible 2D pixels and

vice versa. Ideally, we would like to be able to unproject each pixel into all images accurately which require a perfect depth estimation. Until now, such a multiple image based reconstruction method does not exist; current methods produce inaccurate 3D reconstructed geometry. A Bayesian approach could be used requiring the evaluation of the accuracy of a 3D reconstructed mesh from multiple views, e.g extending the model proposed by [Barron and Malik, 2013a]. Unfortunately, such metrics do not exist; feature matching scores cannot be used since for example a region near shadow boundaries will tend to provide a strong matching score but also noisy normals and inaccurate geometry reconstruction. By opposition, a region without texture has a low matching score because of the lack of texture detail despite being properly reconstructed such as a wall. Apart from multiple views, reconstructed depth from images, sensor or multiple views is a very active field, plausible depth synthesis is emerging as well [Chaurasia *et al.*, 2013], [Karsch *et al.*, 2014a] but none of these approaches are able to characterize the confidence of the reconstructed depth.

Ground truth. Setting up a ground truth evaluation is always time consuming since each intrinsic decomposition method focuses on specific cases from single image, single captured object, multi view stereo reconstructed scene, scanned geometry or even videos. All of these approaches target specific scenes and material appearance models which make it impractical to compare each method to each other efficiently. Until now, the most complete ground truth dataset comes from [Grosse *et al.*, 2009] but focuses on a single object.

7.4 Concluding remarks

This thesis demonstrates extended applications of a ”good enough” intrinsic decomposition algorithm for multi view reconstructed scenes. These applications include relighting and image based content manipulation. The joint development of image based rendering method available on the web (Microsoft Photosynth, Bing Maps, Google Street view etc.), and photogrammetry systems (PMVS, Visual FM, Multi-View Environment, PhotoScan, Autodesk 123D Catch, Smart3DCapture, etc.) promises an increasing interest of 3D scene manipulation methods for architecture, games and the movie industry.

Bibliography

AGARWAL, S., MIERLE, K., and OTHERS. Ceres solver. `http://ceres-solver.org`.

ALLDRIN, N.G. and KRIEGMAN, D., 2007. Toward reconstructing surfaces with arbitrary isotropic reflectance: A stratified photometric stereo approach. In Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, 1–8. IEEE.

BARNES, C., SHECHTMAN, E., FINKELSTEIN, A., and GOLDMAN, D.B., 2009. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 28(3).

BARRON, J. and MALIK, J., 2013a. Shape, illumination, and reflectance from shading. Technical report, Berkeley Tech Report.

BARRON, J.T. and MALIK, J., 2013b. Intrinsic scene properties from a single RGB-D image. *CVPR*.

BARROW, H. and TENENBAUM, J., 1978. Computer vision systems. *Computer vision systems*, 2.

BAYER, B.E., 1976. Color imaging array. *US Patent, 3,971,065*.

BELL, S., BALA, K., and SNAVELY, N., 2014. Intrinsic images in the wild. *ACM Trans. on Graphics (SIGGRAPH)*, 33(4).

BELL, S., UPCHURCH, P., SNAVELY, N., and BALA, K., 2013. OpenSurfaces: A richly annotated catalog of surface appearance. *ACM Trans. on Graphics (SIGGRAPH)*, 32(4).

BLINN, J.F., 1977. Models of light reflection for computer synthesized pictures. In ACM SIGGRAPH Computer Graphics, volume 11, 192–198. ACM.

BONNEEL, N., SUNKAVALLI, K., TOMPKIN, J., SUN, D., PARIS, S., and PFISTER, H., 2014. Interactive intrinsic video editing. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2014)*, 33(6).

BONNEEL, N., VAN DE PANNE, M., PARIS, S., and HEIDRICH, W., 2011. Displacement interpolation using lagrangian mass transport. In ACM Transactions on Graphics (TOG), volume 30, 158. ACM.

BOROS, E. and HAMMER, P.L., 2002. Pseudo-boolean optimization. *Discrete applied mathematics*, 123(1):155–225.

BOUSSEAU, A., PARIS, S., and DURAND, F., 2009. User-assisted intrinsic images. *ACM Trans. Graph.*, 28(5):1–10. ISSN 0730-0301.

BOYKOV, Y., VEKSLER, O., and ZABIH, R., 2001. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11):1222–1239.

BUEHLER, C., BOSSE, M., MCMILLAN, L., GORTLER, S., and COHEN, M., 2001. Unstructured lumigraph rendering. In Proceedings of the 28th annual conference on Computer graphics and interactive techniques, 425–432. ACM.

CARROLL, R., RAMAMOORTHI, R., and AGRAWALA, M., 2011. Illumination decomposition for material recoloring with consistent interreflections. In ACM Transactions on Graphics (TOG), volume 30, 43. ACM.

CHAURASIA, G., DUCHENE, S., SORKINE-HORNUNG, O., and DRETTAKIS, G., 2013. Depth synthesis and local warps for plausible image-based navigation. *ACM Trans. on Graphics (TOG)*, 32(3):30:1–30:12.

CHAURASIA, G., SORKINE, O., and DRETTAKIS, G., 2011. Silhouette-aware warping for image-based rendering. In Computer Graphics Forum, volume 30, 1223–1232. Wiley Online Library.

CHEN, Q. and KOLTUN, V., 2013. A simple model for intrinsic image decomposition with depth cues. In ICCV. IEEE.

COMANICIU, D. and MEER, P., 2002. Mean shift: A robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(5):603–619.

COOK, R.L. and TORRANCE, K.E., 1982. A reflectance model for computer graphics. *ACM Transactions on Graphics (TOG)*, 1(1):7–24.

DEBEVEC, P., TCHOU, C., GARDNER, A., HAWKINS, T., POULLIS, C., STUMPFEL, J., JONES, A., YUN, N., EINARSSON, P., LUNDGREN, T., FAJARDO, M., and MARTINEZ, P., 2004. Estimating surface reflectance properties of a complex scene under captured natural illumination. Technical report, USC Institute for Creative Technologies.

DONG, Y., WANG, J., TONG, X., SNYDER, J., LAN, Y., BEN-EZRA, M., and GUO, B., 2010. Manifold bootstrapping for svbrdf capture. *ACM Trans. Graph.*, 29(4):98:1–98:10. ISSN 0730-0301.

EISEMANN, M., DE DECKER, B., MAGNOR, M., BEKAERT, P., DE AGUIAR, E., AHMED, N., THEOBALT, C., and SELLENT, A., 2008. Floating textures. In Computer Graphics Forum, volume 27, 409–418. Wiley Online Library.

FINLAYSON, G.D., DREW, M.S., and LU, C., 2004. Intrinsic images by entropy minimization. In ECCV, 582–595.

FURUKAWA, Y. and PONCE, J., 2007. Accurate, dense, and robust multi-view stereopsis. In Proc. CVPR. ISSN 1063-6919.

FURUKAWA, Y. and PONCE, J., 2010. Accurate, dense, and robust multi-view stereopsis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 32(8):1362–1376.

GARCES, E., MUNOZ, A., LOPEZ-MORENO, J., and GUTIERREZ, D., 2012. Intrinsic images by clustering. *Computer Graphics Forum (Proc. EGSR)*, 31(4).

GOESELE, M., ACKERMANN, J., FUHRMANN, S., HAUBOLD, C., and KLOWSKY, R., 2010. Ambient point clouds for view interpolation. *ACM Transactions on Graphics (TOG)*, 29(4):95.

GOESELE, M., SNAVELY, N., CURLESS, B., HOPPE, H., and SEITZ, S.M., 2007. Multi-view stereo for community photo collections. In ICCV. ISSN 1550-5499.

GORELICK, L., BOYKOV, Y., VEKSLER, O., AYED, I.B., HEALTHCARE, G., and DELONG, A., 2014. Submodularization for binary pairwise energies.

GROSSE, R., JOHNSON, M.K., ADELSON, E.H., and FREEMAN, W.T., 2009. Ground-truth dataset and baseline evaluations for intrinsic image algorithms. In ICCV.

GUO, R., DAI, Q., and HOIEM, D., 2011. Single-image shadow detection and removal using paired regions. In CVPR, 2011, 2033–2040. IEEE.

GUO, R., DAI, Q., and HOIEM, D., 2012. Paired regions for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99.

HABER, T., FUCHS, C., BEKAERT, P., SEIDEL, H.P., GOESELE, M., and LENSCH, H.P.A., 2009. Relighting objects from image collections.

HORN, B.K., 1989. Obtaining shape from shading information. In Shape from shading, 123–171. MIT press.

JEGELKA, S. and BILMES, J., 2011. Submodularity beyond submodular energies: coupling edges in graph cuts. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, 1897–1904. IEEE.

KARSCH, K., LIU, C., and KANG, S.B., 2014a. Depthtransfer: Depth extraction from video using non-parametric sampling. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*.

KARSCH, K., SUNKAVALLI, K., HADAP, S., CARR, N., JIN, H., FONTE, R., SITTIG, M., and FORSYTH, D., 2014b. Automatic scene inference for 3d object compositing. *ACM Transactions on Graphics (TOG)*, 33(3):32.

KAZHDAN, M.M. and HOPPE, H., 2013. Screened poisson surface reconstruction. *ACM Trans. Graph.*, 32(3):29.

KOLMOGOROV, V., 2006. Convergent tree-reweighted message passing for energy minimization. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(10):1568–1583.

KOLMOGOROV, V. and ZABIN, R., 2004. What energy functions can be minimized via graph cuts? *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(2):147–159.

LAFFONT, P.Y., 2012. Intrinsic Image Decomposition from Multiple Photographs. Ph.D. thesis, University of Nice Sophia-Antipolis.

LAFFONT, P.Y., BOUSSEAU, A., and DRETTAKIS, G., 2013. Rich intrinsic image decomposition of outdoor scenes from multiple views. *IEEE Trans. on Visualization and Computer Graphics*, 19(2):210–224.

LAFFONT, P.Y., BOUSSEAU, A., PARIS, S., DURAND, F., and DRETTAKIS, G., 2012. Coherent intrinsic images from photo collections. *ACM Transactions on Graphics (SIGGRAPH Asia Conference Proceedings)*, 31.

LAFORTUNE, E.P., FOO, S.C., TORRANCE, K.E., and GREENBERG, D.P., 1997. Non-linear approximation of reflectance functions. In Proceedings of the 24th annual conference on Computer graphics and interactive techniques, 117–126. ACM Press/Addison-Wesley Publishing Co.

LALONDE, J.F., EFROS, A.A., and NARASIMHAN, S.G., 2009. Webcam clip art: Appearance and illuminant transfer from time-lapse sequences. *ACM Transactions on Graphics (SIGGRAPH Asia 2009)*, 28(5).

LALONDE, J.F., EFROS, A.A., and NARASIMHAN, S.G., 2010. Detecting ground shadows in outdoor consumer photographs. In European Conference on Computer Vision.

LALONDE, J.F., EFROS, A.A., and NARASIMHAN, S.G., 2012. Estimating the natural illumination conditions from a single outdoor image. *International journal of computer vision*, 98(2):123–145.

LAMBERT, J.H., 1760. Photometria.

LAND, E.H. and MCCANN, J.J., 1971. Lightness and retinex theory. *Journal of the optical society of America*, 61(1).

LEE, K.J., ZHAO, Q., TONG, X., GONG, M., IZADI, S., LEE, S.U., TAN, P., and LIN, S., 2012. Estimation of intrinsic image sequences from image+depth video. In A.W. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, editors, ECCV (6), volume 7577 of *Lecture Notes in Computer Science*, 327–340. Springer. ISBN 978-3-642-33782-6.

LENSCH, H.P.A., KAUTZ, J., GOESELE, M., HEIDRICH, W., and SEIDEL, H.P., 2001. Image-based reconstruction of spatially varying materials. Research Report MPI-I-2001-4-001, Max-Planck-Institut für Informatik, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany.

LENSCH, H.P.A., KAUTZ, J., GOESELE, M., HEIDRICH, W., and SEIDEL, H.P., 2003. Image-based reconstruction of spatial appearance and geometric detail. *ACM Trans. Graph.*, 22(2):234–257.

LEVIN, A., LISCHINSKI, D., and WEISS, Y., 2004. Colorization using optimization. *ACM Transactions on Graphics*, 23:689–694.

LEVIN, A., LISCHINSKI, D., and WEISS, Y., 2008a. A closed-form solution to natural image matting. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(2):228–242.

LEVIN, A., RAV ACHA, A., and LISCHINSKI, D., 2008b. Spectral matting. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(10):1699–1712.

LI, H., FOO, S.C., TORRANCE, K.E., and WESTIN, S.H., 2005. Automated three-axis gonioreflectometer for computer graphics applications. In Optics & Photonics 2005, 58780S–58780S. International Society for Optics and Photonics.

LIAO, Z., ROCK, J., WANG, Y., and FORSYTH, D., 2013. Non-parametric filtering for geometric detail extraction and material representation. In Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, 963–970. IEEE.

LIPSKI, C., KLOSE, F., and MAGNOR, M., 2014. Correspondence and depth-image based rendering: a hybrid approach for free-viewpoint video. *IEEE Trans. Circuits and Systems for Video Technology (T-CSVT)*, 24(6):942–951.

LIU, C., YUEN, J., TORRALBA, A., SIVIC, J., and FREEMAN, W.T., 2008. Sift flow: Dense correspondence across different scenes. In Proceedings of the 10th European Conference on Computer Vision: Part III, ECCV '08, 28–42. Springer-Verlag, Berlin, Heidelberg. ISBN 978-3-540-88689-1.

LOSCOS, C., FRASSON, M.C., DRETTAKIS, G., WALTER, B., GRANIER, X., and POULIN, P., 1999. Interactive virtual relighting and remodeling of real scenes. In Proceedings of the 10th Eurographics conference on Rendering, 329–340.

LOWE, D.G., 2004. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110. ISSN 0920-5691.

MARQUARDT, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics*, 11(2):431–441.

MARSCHNER, S.R., 1998. Inverse rendering for computer graphics. Ph.D. thesis, Cornell University.

MARSCHNER, S.R., WESTIN, S.H., LAFORTUNE, E.P., and TORRANCE, K.E., 2000. Image-based bidirectional reflectance distribution function measurement. *Applied Optics*, 39(16):2592–2600.

MATSUSHITA, Y., LIN, S., KANG, S.B., and SHUM, H.Y., 2004a. Estimating intrinsic images from image sequences with biased illumination. In T. Pajdla and J. Matas, editors, ECCV (2), volume 3022 of *Lecture Notes in Computer Science*, 274–286. Springer. ISBN 3-540-21983-8.

MATSUSHITA, Y., NISHINO, K., IKEUCHI, K., and SAKAUCHI, M., 2004b. Illumination normalization with time-dependent intrinsic images for video surveillance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1336–1347. ISSN 0162-8828.

MATUSIK, W., PFISTER, H., BRAND, M., and MCMILLAN, L., 2003. A data-driven reflectance model. *ACM Transactions on Graphics*, 22(3):759–769.

NICODEMUS, F.E., RICHMOND, J.C., HSIA, J.J., GINSBERG, I.W., and LIMPERIS, T., 1992. Radiometry. chapter Geometrical Considerations and Nomenclature for Reflectance, 94–145. Jones and Bartlett Publishers, Inc., USA. ISBN 0-86720-294-7.

OMER, I. and WERMAN, M., 2004. Color lines: Image specific color representation. In Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, volume 2, II–946. IEEE.

OREN, M. and NAYAR, S.K., 1994. Generalization of lambert's reflectance model. In Proceedings of the 21st annual conference on Computer graphics and interactive techniques, 239–246. ACM.

PANAGOPOULOS, A., WANG, C., SAMARAS, D., and PARAGIOS, N., 2013. Simultaneous cast shadows, illumination and geometry inference using hypergraphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2):437–449. ISSN 0162-8828.

PEARL, J., 1982. Reverend bayes on inference engines: A distributed hierarchical approach. In AAAI, 133–136.

PEREZ, R., SEALS, R., and MICHALSKY, J., 1993. All-weather model for sky luminance distribution – Preliminary configuration and validation. *Solar Energy*, 50(3):235–245.

PERONA, P. and MALIK, J., 1990. Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(7):629–639.

PHONG, B.T., 1975. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317.

PIXAR ANIMATION STUDIOS, Hery, C. and VILLEMIN, R. Physically based lighting at pixar.

PONS, J.P., KERIVEN, R., and FAUGERAS, O., 2007. Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. *International Journal of Computer Vision*, 72(2):179–193.

PREETHAM, A.J., SHIRLEY, P., and SMITS, B., 1999. A practical analytic model for daylight. In SIGGRAPH, 91–100.

RAMAMOORTHI, R. and HANRAHAN, P., 2001. A Signal-Processing Framework for Inverse Rendering. *SIGGRAPH*, 117–128.

ROMEIRO, F., VASILYEV, Y., and ZICKLER, T., 2008. Passive reflectometry. In Computer Vision–ECCV 2008, 859–872. Springer.

ROMEIRO, F. and ZICKLER, T., 2010a. Blind reflectometry. In Computer Vision–ECCV 2010, 45–58. Springer.

ROMEIRO, F. and ZICKLER, T., 2010b. Inferring reflectance under real-world illumination. Technical report, Technical Report No. TR-10-10). Cambridge, MA: Harvard School of Engineering and Applied Sciences.

ROTHER, C., KOLMOGOROV, V., and BLAKE, A., 2004. Grabcut: Interactive foreground extraction using iterated graph cuts. In ACM Transactions on Graphics (TOG), volume 23, 309–314. ACM.

SAVARESE, S., ANDREETTO, M., RUSHMEIER, H., BERNARDINI, F., and PERONA, P., 2007. 3d reconstruction by shadow carving: Theory and practical evaluation. *International journal of computer vision*, 71(3):305–336.

SHEN, J., YANG, X., JIA, Y., and LI, X., 2011. Intrinsic images using optimization. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, 3481–3487. IEEE.

SHEN, L., TAN, P., and LIN, S., 2008. Intrinsic image decomposition with non-local texture cues. In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 1–7. IEEE.

SHIH, Y.C., PARIS, S., DURAND, F., and FREEMAN, W.T., 2013. Data-driven hallucination of different times of day from a single outdoor photo. *ACM Trans. Graph.*, 32(6):200.

SNAVELY, N., SEITZ, S.M., and SZELISKI, R., 2006. Photo tourism: exploring photo collections in 3d. In ACM transactions on graphics (TOG), volume 25, 835–846. ACM.

SOILLE, P., 2003. Morphological Image Analysis: Principles and Applications. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2 edition. ISBN 3540429883.

STARK, M.M., ARVO, J., and SMITS, B., 2005. Barycentric parameterizations for isotropic brdfs. *Visualization and Computer Graphics, IEEE Transactions on*, 11(2):126–138.

SUNKAVALLI, K., MATUSIK, W., PFISTER, H., and RUSINKIEWICZ, S., 2007. Factored time-lapse video. *ACM Transactions on Graphics (proc. of SIGGRAPH)*, 26(3). ISSN 0730-0301.

SUNKAVALLI, K., ROMEIRO, F., MATUSIK, W., ZICKLER, T., and PFISTER, H., 2008. What do color changes reveal about an outdoor scene? In CVPR. IEEE Computer Society.

SZELISKI, R., 2010a. Computer Vision: Algorithms and Applications. Springer-Verlag New York, Inc., New York, NY, USA, 1st edition. ISBN 1848829345, 9781848829343.

SZELISKI, R., 2010b. Computer vision: algorithms and applications. Springer.

TAO, L., YUAN, L., and SUN, J., 2009. Skyfinder: attribute-based sky image search. In ACM Transactions on Graphics (TOG), volume 28, 68. ACM.

TAPPEN, M.F., ADELSON, E.H., and FREEMAN, W.T., 2006. Estimating intrinsic component images using non-linear regression. In Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, volume 2, 1992–1999. IEEE.

TAPPEN, M.F., FREEMAN, W.T., and ADELSON, E.H., 2005. Recovering intrinsic images from a single image. *IEEE Trans. PAMI*, 27(9).

VEACH, E., 1997. Robust Monte Carlo methods for light transport simulation. Ph.D. thesis, Stanford University.

WAINWRIGHT, M.J., JAAKKOLA, T.S., and WILLSKY, A.S., 2003. Tree-reweighted belief propagation algorithms and approximate ml estimation by pseudo-moment matching. In Workshop on Artificial Intelligence and Statistics, volume 21, 97. Society for Artificial Intelligence and Statistics Np.

WEISS, Y., 2000. Correctness of local probability propagation in graphical models with loops. *Neural computation*, 12(1):1–41.

WEISS, Y., 2001. Deriving intrinsic images from image sequences. In ICCV, 68–75. ISBN 0-7695-1143-0.

WEISTROFFER, R.P., WALCOTT, K.R., HUMPHREYS, G., and LAWRENCE, J., 2007. Efficient basis decomposition for scattered reflectance data. In J. Kautz and S.N. Pattanaik, editors, Rendering Techniques, 207–218. Eurographics Association. ISBN 978-3-905673-52-4.

XIAO, J., HAYS, J., EHINGER, K.A., OLIVA, A., and TORRALBA, A., 2010. Sun database: Large-scale scene recognition from abbey to zoo. In Computer vision and pattern recognition (CVPR), 2010 IEEE conference on, 3485–3492. IEEE.

YE, G., GARCES, E., LIU, Y., DAI, Q., and GUTIERREZ, D., 2014. Intrinsic video and applications. *ACM Transactions on Graphics (SIGGRAPH 2014)*, 33(4).

YU, Y., DEBEVEC, P., MALIK, J., and HAWKINS, T., 1999. Inverse global illumination: Recovering reflectance models of real scenes from photographs. In Proceedings SIGGRAPH'99, 215–224.

ZHAO, Q., TAN, P., DAI, Q., SHEN, L., WU, E., and LIN, S., 2012. A closed-form solution to retinex with nonlocal texture constraints. *IEEE Trans. PAMI*, 34:1437–1444. ISSN 0162-8828.

ZHU, J., SAMUEL, K.G.G., MASOOD, S., and TAPPEN, M.F., 2010. Learning to recognize shadows in monochromatic natural images. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2010).